VOLUME X APRIL, 1919 NUMBER 4

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

THE RESPONSE OF A COMPOSITE GROUP TO THE STANFORD REVISION OF THE BINET-SIMON TESTS

H. W. CHASE AND C. C. CARPENTER
University of North Carolina

During the winter and spring of 1916–1917, the writers tested 103 children in the elementary school of Chapel Hill, N. C., by the Stanford Revision of the Binet Tests. After some preliminary training work, the actual testing was done by one of us (C. C. C.). Terman's directions, as laid down in his *Measurement of Intelligence*, were followed absolutely, and conditions were favorable in all ways throughout.

For the purposes of this study, the records of all children below 9 years 0 months and above 12 years 11 months were disregarded. This leaves a total of 77 records, comprising those of all children in the school whose chronological ages ranged from 9 to 12 years inclusive. Most of these children were in the fourth, fifth, and sixth grades.

The relevant facts as to the distribution of the intelligence quotients of this group of 77 children are shown in Table I.

It will be seen that our distribution is somewhat skewed. Our results differ throughout from Terman's in that the median mental and median chronological ages fail to coincide at any of our age-levels. For children of the ages tested by us, Terman reports median I. Q.'s as follows: 9 years, 100.5; 10 years, 103; 11 years, 98; 12 years, 98.[1] The average of his medians is, for the four years under consideration, 99.9. Our medians for these ages have been stated in Table I; their average is 92.2 and only 23% of our children reach or exceed the average of his four medians. The average of the medians of our two larger age-groups (10 and 12 years) is 91.9.

The divergence of our results from Terman's is too considerable

[1] *The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence.*

to be dismissed as due to possible variations in our technique from that prescribed by Terman. It is to be explained rather by the

TABLE I

DISTRIBUTION OF I. Q. OF 77 CHILDREN FROM 9 TO 12 YEARS

	Age 9	Age 10	Age 11	Age 12	All
No. tested	11	26	12	28	77
Range of I. Q.	90.4–118.4	68.8–140.7	69.6–119.2	58–123.8	58–140.7
Middle 50%	94.3– 99.9	88.1-103.2	75.3– 98.2	83.2–97.3	85.2–99.7
Median	98.2	93.1	86.7	90.7	92.2
No. with I. Q. below 70	0	1	1	1	3
No. between 70-80	0	0	3	5	8
No. between 80-90	0	10	3	7	20
No. between 90-100	7	7	2	12	28
No. between 100-110	3	5	1	0	9
No. between 110-120	1	1	2	2	6
No. between 120-140	0	1	0	1	2
No. above 140	0	1	0	0	1

composition of the group itself. Of the 77 children, 12 are members of families connected professionally with the University of North Carolina. The intelligence quotients of this group of children are 100, 103.2, 106.3, 106.5, 110.6, 112.1, 118.4, 119.0, 119.2, 120.3, 123.8, 140.7. The median is 115.25. All the children, it will be noted, are above the average of Terman's medians for the indicated years. In terms of his interpretation of the significance of intelligence quotients,[2] four of these children are of average intelligence (90–110), five of superior intelligence (110–120) and three of very superior intelligence (120–140); one of the three, indeed (140.7) reaching a level attained by only one child in 250 or 300.

The remaining 65 children form a rather homogeneous group, descended in the great majority of cases from stock which has lived for generations in this or in closely similar localities. Many of the families are inter-related, and all have in general the same standards, traditions, and a common range of experiences and ideas. The fathers are small business men, skilled workmen, etc., and small farmers living in the town or in the adjacent country districts. Thirteen of this group live outside the town limits, and had attended

[2] *The Measurement of Intelligence*, p. 79.

the school at the time when the tests were made for a median time of two years. They are thus somewhat less favored educationally than the town group, as the country schools from which they came are without exception poor. The intelligence quotients of this "country" group of 13 are 69.6, 73.1, 75.3, 75.6, 76, 77.7, 81.5, 88.1, 88.2, 88.3, 89.2, 94.2, 96.6, with the median at 81.5. No one of these children, it will be noted, reaches 100. The faculty" and the "country" groups are thus completely separated by the tests, the median of the latter group being 33.5 below that of the former.[3]

Of the remaining 52 (the "town" children), the median I. Q. is 92.3. Two have I. Q.'s of less than 70 (58, 68). Four exceed 100, the highest record being 111.4 (mother a teacher).

On the basis of the data given by Terman in Appendix 1 of the "Stanford Revision," the attempt was made to discover whether the deficiency of the 65 town and country children was general, or was localized in certain tests. The data obtained by Termna, Lyman, Ordahl, etc., by Terman and Childs, and by Terman, Trots and Waddle, were utilized in the attempt to discover how the percentage of children of a given age passing a given test compared in Terman's investigations and in our study. As our age-groups are small, and as the percentages of children of a given age passing a given test vary, sometimes considerably, in the different investigations reported by Terman, any difference between Terman's figures and our own for children of any one age in any one test naturally means little or nothing. It may be assumed, however, that any marked deficiency in our children of all four ages in any one test will possess some significance. Such deficiencies exist in tests 10–1, 10–3, 10–6, 12–1, and 12–2. In other words, children of all four ages tested by us were deficient in vocabulary (10–1, 12–1), in drawing designs from memory (10–3), in the giving of sixty words in three minutes (10–6), and in the ability to define abstract words (12–2). The deficiency in 10–3 (drawing designs) is possibly due to the fact that instruction in drawing is not given in the school. The other deficiencies suggest at least that the children's stock of speech-habits is small, and that those which have been acquired are by no means in a condition of readiness to function. The results of practically all tests, however, were poor. In only two cases (making change and reading and report), were the scores of all our age-groups

[3]Eight other country children, not within the designated ages, were tested in the course of the work. Their I. Q. 's are 48.3, 53.4, 64.3, 67, 71.8, 75, 82, 83.7. No other faculty children were tested.

up to Terman's Standards. In view of this and of the facts noted above, it is doubtful whether much importance should be attached to the particularly bad records in the tests above mentioned. The records of the faculty children in these tests are good. In naming 60 words all passed except one child of three at nine years and one of four at ten years. At nine and ten, all passed the 10-year vocabulary, one at nine and two at ten passing the 12-year vocabulary also. At eleven and twelve, all passed the vocabulary tests for 12 and 14 years.

It is interesting to note that the reaction of our town and country group of 65 to the tests is considerably more homogeneous than that of the children of the same ages tested by Terman. The average deviation of the scores of the 362 children from 9 to 12 tested by him (see pp. 35 and 36 of the *Stanford Revision*) is approximately 10, ours is 7.5. This greater homogeneity is due largely to the fact that the I.. Q.'s of 43.1% of the 65 children were bunched between 86–95, while only 33.9% of Terman's children are included within any 10–point range (that from 96–105). The percentage distribution of the I. Q.'s of these 65 children, as compared with Terman's distribution, is shown in Table 2. The greater homogeneity of our group is natural in view both of the interrelated stock from which the children come and of the similarity of their environment.

TABLE II

PERCENTAGE DISTRIBUTION OF I. Q.'S FOR 65 CHILDREN (CHAPEL HILL) COMPARED WITH DISTRIBUTION FOUND BY TERMAN

I. Q.	Terman (%)	Chapel Hill (%)
56–65	0.33	1.5
66–75	2.3	7.7
76–85	8.6	20.0
86–95	20.1	43.1
96–105	33.9	23.1
106–115	23.1	4.6
116–125	9.0	0.0
126–135	2.3	0.0
136–145	0.55	0.0

It is clear that we have in our group of 77 children three different levels of response to the test situations. The high-level response of the faculty group is in agreement with the findings of other in-

The Terman percentages aref rom the *Measurement of Intelligence*, p. 66. They are his combined results for children from 5-14. These figures differ so slightly from those for children for 9-12 years that it is a matter of indifference which data are used for the comparison.

vestigators who have tested children of superior social status. The lower level of response of the town and country group is more difficult to explain. It is certainly not due to the fact that these groups are as a whole, or to any considerable degree, made up of children of "inferior" social status in the sense in which Terman and others have used the term. Judged by the standards of small-town communities throughout the state, most of the town children are clearly of average social status, and the same is true of the country children if judged by the prevailing standards of rural life. It is certainly not the case that the thirteen country children between the ages of 9 and 12, and the eight others without those ages, come from the inferior homes of their neighborhoods. It is not the children of the most ignorant and unprogressive farmers who are sent into town to school. And yet, of these 21 children, not one has an intelligence quotient reaching 100, and eight have intelligence quotients of 75 or less. Of the 52 town children, very few indeed are from homes sufficiently below the average of the typical small-town community to be called inferior, while a few are from homes distinctly above the average for their group. Yet the median intelligence of the group is 7 points lower than that of Terman's average group. (*Stanford Revision*, p. 91)

It may seem that to emphasize the fact that these children come from "average" homes is simply to quibble over words; that the only fact worth considering is that the home environment of those children, though average as measured in terms of their own fellows, may be inferior if judged by the standards of environment in which Terman's work was done. The case is, however, by no means so simple. From the average homes of any community, whatever measured in absolute terms their standards may be, come the judgments and opinions which largely regulate the community life. The inferior home, however great its social importance may be in other ways, is by no means equally important in shaping community standards and activities. Thus the social importance of a low or high level of intelligence in the average home is obviously very great. We are far from intimating that the intelligence quotients of children from average homes in the small towns and rural neighborhoods of this State, or of this section of the State, average lower than those of the Western children examined by Terman. We have shown only that such a condition does exist in children now in the upper grades of the elementary school in this community, excluding the "faculty" children. From the results of scattering tests made

in other grades, and from our observation generally, it seems probable that the condition exists throughout the school.

Our work has, however, gone far enough to convince us that, if the Stanford or any other intelligence scale is to be made available for use throughout the country, a great amount of work needs to be done in the derivation of norms. It seems to us probable that no general norms for children of inferior, average, and superior social status can be set up, but that sectional norms, perhaps norms for states or for homogeneous groups of states, will be needed. Furthermore, most measurements of intelligence have been made with city children. Can rural children be judged by the same norms? Or can they in some sections of the country and not in others? Is it possible to judge satisfactorily the intelligence of an individual without knowing first what normal intelligence means in the community from which he comes? For so long as he remains in that community, his successes and failures will be determined, not by any abstract standard of "average" intelligence, but by the relation of his mental ability to that of his neighbors.

Such questions become of great importance—to take but one illustration—in any attempt to use intelligence scales to diagnose feeble-mindedness. Terman, though recognizing that the boundary between normality and subnormality is "absolutely arbitrary," states definitely that "all who test below 70 I. Q. by the Stanford revision of the Binet-Simon scale should be considered feeble-minded, and it is an open question whether it would not be justifiable to consider 75 I. Q. as the lower limit of 'normal' intelligence."[5] Now of the 21 country children tested in all by us, 5 test below 70 I. Q. and 7 below 75 I. Q. Are these children all to be classed as feeble-minded? Two of them are clearly such. We seriously doubt whether the others can be so classified. They are not, to be sure, capable of progressing far in school; but, on the other hand, there is every reason to believe that, in the simple environment in which they will pass their lives, they will sustain themselves satisfactorily. When it is remembered that the median intelligence quotient of the country group is approximately 80, it is clear that individuals who test between 60 and 70 will be able to compete on terms which are not too unequal.

It is possible that we shall find that we need to distinguish between absolute and relative feeble-mindedness. A judgment of absolute feeble-mindedness would mean, say, that the individual

[5] *The Measurement of Intelligence*, p 81.

tested below some point set up on the basis of nation-wide or even inter-national experimentation with intelligence scales. It would signify a lack of that degree of intelligence necessary to enable the individual to adjust himself satisfactorily to the demands of life in typically "modern" communities, with their complexity of organization. Such absolute standards are of course the ones which need to be considered in such occupations as army service, in which the demands made on the intelligence of men are the same throughout the country. A judgment of relative feeble-mindedness, on the other hand, would mean that the individual was incapable of adjusting himself to the demands of life in his own, or in similar environment. The point below which it would lie would vary with many factors. It is, of course, possible to carry the argument to an absurd extreme, and to contend that such individuals as the Jukes should be considered "relatively" normal, since they get on well enough at their low level of existence. But such individuals are considered, even by their neighbors in the same general environment, as abnormal. The point is just that such individuals cannot adjust themselves satisfactorily to the life of even the simplest sort of decent and self-respecting modern civilized community. It is altogether probable, on the other hand, that individuals with intelligence quotients of between 60 and 70 *can* adjust themselves fairly satisfactorily to the demands which life makes on the average small farmer in the vicinity in which our tests were given. To the more complex life of a larger community such individuals would probably fail to adjust themselves; in this sense their normality or feeble-mindedness is relative.

Intelligence tests given to Southern negroes, again, will unquestionably show that, judged by an absolute standard, enormously high percentages are feeble-minded. But certainly one would miss the point altogether who should fail to remember that, in terms of their own environment and race, most of these individuals are normal enough.

But to return to the tests. The median I. Q. of the faculty group is, it will be remembered, 115, that of the town group is 92.3. This difference of 23.7 points is considerably greater than Terman found between his inferior and superior groups (14 points).[6] For the ages tested, it amounts to a difference of from two to three years in mental age. It is a difference as great as that found by any investigators between groups of children of unlike social status.

[6] *The Stanford Revision*, p. 91.

Thus, Yerkes, who tested children of an inferior and of a superior group in the Cambridge schools, estimates "the superiority of the favored individual over the unfavored as 20 per cent." by the point scale, and adds that this difference is doubtless too low for middle childhood and too high for early adolescence.[7] In terms of mental age, this means a difference, at the ages tested by us, of from two to three years. With this result the work of Bridges and Coler in Columbus closely agrees.[8] It seems fairly certain, then, that American children of the ages tested by us may differ, because of unlike social status, by from two to three years in mental advancement.

Now does unlikeness in social status produce differences in the level of response to intelligence tests predominately because of unlike native endowment in the children tested or because of unlike environments in which they have lived? Yerkes, and Bridges and Coler, prefer not to raise the question, while Terman concludes that native endowment is the predominant factor.

It is clear that the favored groups, both in Cambridge and in Columbus, were from environments different in all important respects from those of the unfavored groups. Not only did the two sorts of children in each city come from homes of different type, but they had attended different schools, their free time had been spent with different companions, the neighborhood influences as a whole were different. The unfavored children tested by Yerkes, though all from English-speaking homes, lived in a quarter largely populated by foreigners. Bridges and Coler mention specifically the poor school building, inferior teachers, and bad neighborhood influences generally of their unfavored group.

Now faculty and town children tested by us differ far less from each other in their general environment than do the superior and inferior Cambridge and Columbus groups. Of the faculty children, seven were born in the town and all the remainder save one—who came from a closely similar environment—have lived in the town since before their school-days began. The two groups have attended the same school, learned from the same teachers, played together freely in and out of school, been cared for in their early years by the same type of colored servants, enjoyed the same pastimes, reacted to the same rural environment. In home life and in occasional travel only has the faculty group had a marked advantage.

[7] YERKES, R. M. *A Point Scale for Measuring Ability*, Chap. 6.

[8] BRIDGES, JAMES W., AND COLER, LILLIAN F. *The Relation of Intelligence to Social Status.* Psy. Rev. 1917, *24*, 1-31.

In many important features the environment of the two groups has been practically the same. Yet they differ in their test records as greatly as do children from good and poor quarters of cities, who have little in common in the way of environment. That such differences should still exist when the action of unlike environmental influences is reduced as nearly to a minimum as it could well be in children of different social status suggests strongly that the records of individuals in tests like those making up the intelligence scales now in use are determined predominately by their native endowment. It is difficult to believe that the homes from which the faculty group comes are so superior in cultural influences to those of favored groups tested in other places as to compensate for the absence of other environmental differences.

We would not, of course assert that environmental influences have nothing to do with the results. On the contrary, the poor record of the unfavored children in such language tests as those of vocabulary and sixty words is clearly in part due to the fact that little reading goes on in the homes from which many of them come, and that outside of school they come less into contact with the world of ideas. But it should not be forgotten that the reaction of the unfavored children to the tests as a whole was poor; it would be difficult, for example, to explain on the ground of environment the fact that children of two of the four ages tested (10 and 11) were notably deficient in the ball and field test—surely if environment ever favors such a test, it does so in a small-town community.

If heredity is the predominant factor in the individual's reaction to the tests, the unusually good record of the faculty group is only to be expected, since the families from which they come owe their presence in the town to the fact that at least the father has been chosen as a member of the community because of excellence in mental qualities presumably closely similar to those tested by the intelligence-scale.

It might seem that the hypothesis that heredity is the main determining factor in the responses breaks down when one attempts to explain by it the difference between the town and country groups. These two groups are of homogeneous stock, and much intermarriage has taken place between their families. Differences in environmental influences do exist; the standards of living are somewhat lower in the country group, they have attended the town school for shorter periods and suffered from inferior teachers in their early school days, their environment as a whole has been less stimulating,

their attendance at school has been more irregular, and conditions as to health and sanitation in their neighborhoods have not been so good.

But, to one who is familiar with the locality, the influence of an heredity factor is clearly apparent. The farming land about the village is generally poor, and the rewards of the average small farmer's labor have been as a rule meager. The incentives for intelligent and ambitious individuals to remain in the locality have thus been slight. That there has been a drift of population away from the county, which contains only two towns, both of less than 2000 inhabitants, is evident from the fact that the population of the county, which was 14,948 in 1890, was only 15,064 in 1910. This drift has been both to the towns and to cities without the county or the state. Thus, there was, in the 20 years just mentioned, an increase in the village itself of 132. As the whole township of which the village forms part only increased in the same period by 80, it is clear that there was in the rural sections of the township an actual decrease in population during the 20 years. Some of this decrease is accounted for by migrations of more ambitious individuals to the village in search of better opportunities, part of it by migration to other localities for the same reason. The country neighborhoods have, in other words, been subject to an adverse selection, which has undoubtedly removed many of the more capable individuals. The poor environmental influences of the children would then, in large measure, have to be considered as symptoms rather than as causes. Though the town has profited to some extent by immigration from the country districts, it has, because of its rather restricted opportunities, been subject to the same adverse selection. Its small net increase in population during the 20 years up to 1910 has been mentioned. The fact that the faculty families have been attracted *to* the town because of intelligence above the average, while the more able "native" inhabitants of the town have been attracted *away* from it for the same reason, has then possibly produced the situation revealed by the tests. In this fact, rather than environmental differences, lies the explanation of the unlike responses of the groups.

It remains only to suggest that many thousands of small-town and rural communities throughout the country have been subjected to this process of adverse selection. What has been the effect of this selection upon the capacity of the younger generation in such functions as those tested by the scale. The answer to this question would be of the very greatest social and educational importance.

AN EIGHTH GRADE DEMONSTRATION CLASS AND THE THREE R'S

COLIN A. SCOTT
Mount Holyoke College

The village school of South Hadley, Massachusetts, was brought, in 1917, into connection with the Education Department of Mount Holyoke College by a small grant from the college which was used in paying part of the 8th grade teacher's salary. The purpose was to make a school which would be able to demonstrate some of the principles of modern education and afford a means of observation and study for the students of the Education Department. It was thought best for the first year to concentrate effort on one grade and to demonstrate there what could be done by practical scientific analysis followed up by such methods as could be based on this analysis.

There was no attempt to suggest a system of preconceived methods for the teacher. Even the most philosophically founded principles of education are liable to prove abortive when handed over to the teacher in the form of methods. In the demonstration class, on the contrary, we were accustomed to say, "When success flatters or when difficulties arise, analyze the situation. Look for your methods afterwards." If this itself may be called a method, it is one of attitude. It does not consist of rules previously learned.

ARITHMETIC

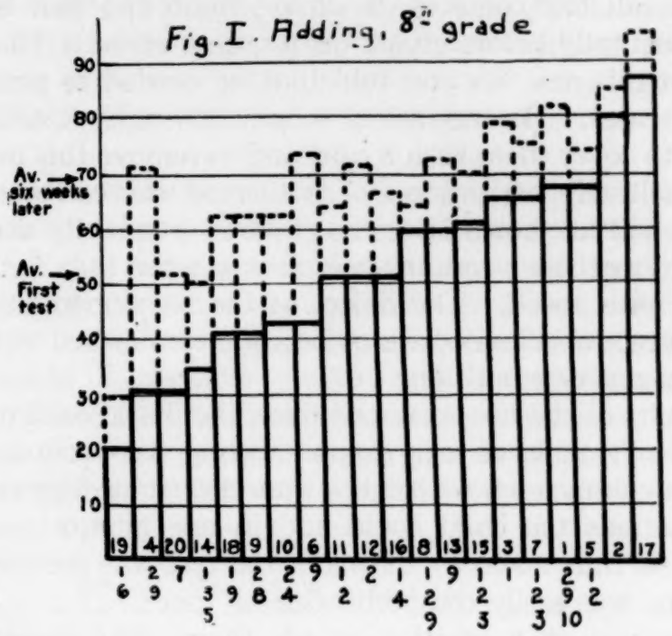
The situation in this 8th grade demonstration class which first attracted us was that of arithmetic. For administrative reasons, not necessary to detail, a beginning was not made until late in November. At this time the children were doing seat-work on Courtis blanks for adding, subtracting, multiplying and dividing, and were having some oral drill in adding, etc., from tables on the blackboard. This was done in adding, for example, by having one child at a time add while the other children were supposed to follow. It is obvious that if the child chosen is a rapid adder, the other children are unable to follow, while if he is slow, the others can hardly be interested enough to keep pace with him, or if they do the practice is useless to them. One person besides the teacher does the work. If the class is twenty in number, therefore, it ought to be

possible to find a way to have twenty times as much work done in the time. A situation very like this is found in many other subjects beside arithmetic where the traditional class teaching is the ideal pursued. As for the seat work on the Courtis blanks, it was found that although all the children were working, the practice did not seem to improve their speed to any extent. The difficulty standing in the way of the progress of each child had not been discovered. Under these conditions, mere repetition may result in perpetuating a habit essentially slow.

Besides the drill work in the four simple operations, work in problems was also going on. This was of the conventional character drawn from the book and consisted in offering the children problems to solve on the analogy of problems previously explained by the teacher. There was no opportunity provided for the children to frame problems for themselves on the basis of their observation and experience of things with which they came in contact. There was no stimulus therefore for them to inquire along quantitative lines. They did not ask questions. They merely answered them, and this naturally according to stereotyped models previously provided. Procedure of this kind is traditional and widespread, but since the inventive and inquiring attitude is absent, there is very little about it which can be said to exercise the reasoning powers.

The more careful and scientific analysis of the situation in arithmetic was begun upon the four simple operations, the other work going on meanwhile, as it had always been conducted. To economize space we will follow the analysis of adding, and omit the full description of the work done on the other three simple operations; but the reader will understand that similar work was done in these subjects also.

One reason for choosing the four simple operations to start with was that in the case of adding, for example, the work to be done to make oneself an efficient adder is simple and limited. It is simple because the basis of the operation requires only the immediate association of a pair of digits, given as the stimulus, to the digits representing their sum, with a background consciousness of the set, attitude, or task to be accomplished. The last factor is not one of great educational importance or difficulty, but it is nevertheless true that children in adding a list of digit pairs will sometimes give the multiplied product instead of the sum of the numbers before them. The task itself is simple. To see, for example, seven and nine, to read



these digits thus, pronouncing them mentally, and immediately to respond by saying, either aloud or mentally, the word sixteen. It is not needed and it is not desirable to think of these signs as numbers, and it is very undesirable to get at the word sixteen by any process of calculation. The immediate response is most efficient and characterizes every good adder. It is obviously a matter of verbal association and no sense of further significance need clog the mind of the learner in the middle and upper grades at least.

Besides being simple, the material in adding (as in the other three simple operations) is limited. The combinations or associations required are not numerous. The combinations from one to ten are the square of ten or one hundred. If we take the digits up to nine, the square of this is 81, and this includes combinations which are the same when reversed.

Printed lists in columns were prepared, containing these combinations arranged in horizontal pairs and in random order. The children were tested by asking each by himself to add these pairs as rapidly as possible; the teacher who tested them being provided with a stop watch, and a similar list but with the answers put opposite each pair. The teacher's attention was thus free to check any pair incorrectly added and any pair in which noticeable delay occurred. These delayed or wrong combinations troubled certain

children and it was also found that often the cause of delay was that a wrong result had come to the child's mind and that he had corrected it mentally before giving his response aloud. The child was then given this new list and told that he needed to practice these particular cases. He was asked to put the right answers opposite these but to cover them with a card and to remove this immediately when he felt any hesitation, so as to read the correct answer at once. As will be shown later the children frequently use incorrect roundabout methods of adding, which even when they were accurate interfered with speed. The object of the recommendation was to get the correct and immediate association established in the child's mind by means of repetition.

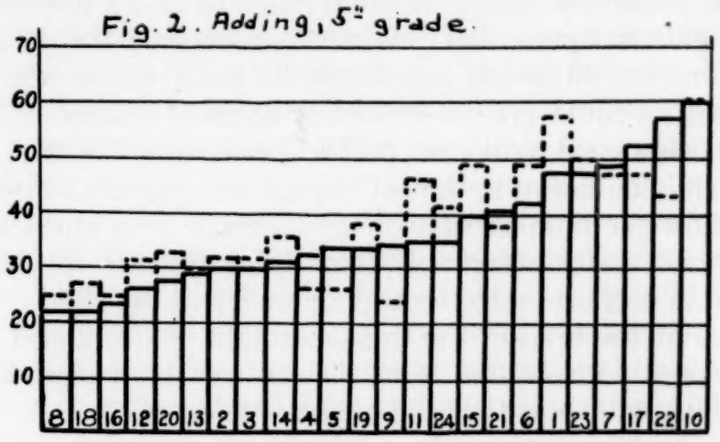
The results of the test were put upon the blackboard in the form of a frequency table or in a graph showing the attainment of the children in columns whose heights were determined by the number of combinations the child could add in one minute (see Fig. 1). The analysis thus made by means of the test was presented to the children and was easily comprehended by them.

We were now in a position to ask them what should be done about it. Those who were low in adding wished to improve. Those who were high wished to know what standard should be attained. We adopted the standard of sixty combinations in a minute for the 8th grade and told the children that it was not necessary for them to practice further after they had reached this point, and that their time could be put on something else. Many children, however, practiced after they had reached this point. Never more than five minutes a day was allowed in school for practice.

The method proposed was that each child should get some other child to hear him and check his delays and inaccuracies when he practiced, just as the teacher had done in the test. When any child thought he had progressed to an appreciable extent, he came to the teacher for a more official test, and was moved up on the frequency table. The children who needed practice often came to school early in order to carry this on. At home they also did some practice, getting a relative to hold the check list for them. They were told that they would probably get better results in a shorter time if they practiced five minutes at a time and repeated this several times. They were asked not to practice more than twenty minutes a day, in order not to interfere with other studies. The carrying out of the work was left entirely in their own hands.

As will be noticed the children used the tests themselves and after the first analysis had been made, pursued methods which were understood by them and were partly their own invention, and even when suggested by the teacher were always under their own choice and control.

On January the seventh, all the children were tested by the techer with the results shown in the graph. All but three of the children had arrived at the sixty combinations a minute standard. Number 19 was out of school most of the time through illness and the Christmas holidays have intervened. Omitting the holidays this leaves about six weeks of practice. As the graph shows the average had moved up from 52 to 68. Judging from the results obtained in the 5th grade of the same school, this advance is as much as had been made in three previous years of their school attendance (see Fig. 2).



It will be noticed on comparison of the first 8th grade test with that of the 5th grade, that the variation is greater in the 8th grade. It would seem probable that the reason for this is somewhat as follows. In the 5th grade the children are near the point where adding (and the three other simple operations) have been emphasized as such. From this time on the work in arithmetic would assume adding to have been learned. The children would then get only the incidental practice which came from the solving of problems. In doing these many children were doubtless slow and would not get through when the rest of the class had already finished.

These pupils, would, moreover, make more mistakes in adding. As a consequence the pressure on their attention because of their difficulties in adding would prevent them giving their best attention to the nature of the problem itself. The result would be that both these children and the teacher would be disposed to think that they were poor in arithmetic and could not readily understand the problems. The children would become discouraged, would do less work, and would get less incidental practice in adding itself. This would account for the children at the lower end of the class who had not improved beyond the average of the 5th grade.

A further analysis was made in order to discover the roundabout methods used by the children. The adding list was put in a machine which exposed one pair of digits at a time and which could be regulated so as to run at the average rate at which the child was able to add. Those combinations therefore which fell below the average were thus picked out, since the child failed to give a response before the next pair was presented. When this occurred the machine was stopped and the child was asked how he had tried to add. It was found that he could give a very clear account of this. The small numbers below each individual on the graph show the kind of errors for each child as shown in the list of such errors given below. They do not show the number of such errors made in a given time. It was however easily observable that the frequency of these errors decreased in proportion to the adding ability of the child. The two who stand at the head of the frequency table (Nos. 2 and 17) were the only ones entirely free from these roundabout methods. The following list will show the kind of errors referred to.

1. (a) Doubling one number and adding or subtracting 1. Nine persons.
   Combinations affected, 7+8; 9+8; 6+5.
   (b) Doubling one number and adding or subtracting a number larger than 1.
   Combinations affected, 6+8; 8+5.
2. Using a familiar combination and adding or subtracting 1. Ten persons. Combinations affected, adding 1. 9+4 (9+3); 8+5 (8+4). Subtracting 9+5 (10+5); 9+4 (9+5).
3. Adding to make 10. Three persons.
   Combinations affected 7+9; 6+9; 7+4.

4. In combinations of 8 or 9 subtracting 2 or 1 respectively from the smaller numbers to get the last digit of the result. Three persons.
   Combinations affected, 9+6; 8+5.
5. In combinations of 9 adding 10 and subtracting 1. One person.
   Combinations affected, 8+9.
6. Adding by 2's. One person.
   Combinations affected 8+4.
7. Auditory memory, *i. e.*, had to whisper combinations before giving them. One person.
8. Finger adding, actually or by thinking of fingers. One person.
9. Learned wrong result. Five persons.

It will be noticed that the graph for the 5th grade represents the record of two tests, both of which were taken about the same time, and with no special practice between. This is for the purpose of determining the reliability of the tests. If two tests are taken on or about the same day, the relative positions of the children in these two tests ought to correspond rather closely if the test is reliable. If the children are not used to the test, there will also be a slight improvement, on the average, in the second test. But variations occur either on the part of the teacher who gives the test or on the part of the child, which will make the record of one test a little different from the record of another. If the correspondence in relative position of all the children (or correlation as it is technically called) is perfect, it is spoken of as 100%, and from this it falls off until a mere chance correspondence obtains. The amount of correlation is easily calculated and in this case was found to be 85%, therefore representing a high degree of reliability on the average for the test in adding.

As we have already said, similar work was done on the other three simple operations. On comparing the correspondence of standing of the pupils, or the correlation between the first set of tests and the last tests after systematic practice had intervened, it was found that the correlations had largely increased. The correlation between adding and subtracting, for example, before practice was 64%, and after practice 92% or higher than for two tests in adding. This means that there must have been children in the first tests who had not reached their natural place in one subject or the other and that

practice had given them a standing more nearly corresponding to their natural capacity for learning. It was found also that the correlation between silent reading and adding, although not high (15%), had increased considerably (to 30%) after practice in adding.

As suggested earlier, adding is really quite like reading. The child actually reads the figures and responds by work pronounced mentally or spoken aloud. Reading also involves habitual associations as in adding; as a result of which conceptions arise in the mind, having a mental sign, sometimes a word, not always in the text itself. One reads faster when the style is such that the words in the first part of a sentence have stimulated associations represented by words in the latter part of the same sentence. The same thing is true in passing from one sentence to another, and for that part of it the feeling of the continuity of the theme itself is caused by an expectation being aroused which is later on satisfactorily fulfilled. It is not too much to say then, that adding (and indeed mathematics more complicated than the simple processes) may be regarded as a kind of reading. In support of this we may observe that one operation in adding must take the use of three words, although these in the case of fast adders may not be fully pronounced. Some experienced adders can add about 230 combinations in a minute. This would mean about 700 short words per minute, and I have known readers who were capable of reading (silently)easy texts containing many short words at this rate. In the 8th grade graph the best adder (No. 17) adds about 90 combinations a minute. At the rate of three words per operation, this would give 270 words per minute, which is about the rate at which he reads easy texts. No. 19 is a slow reader (about 150 words a minute), No. 4, however, is a much faster reader and this probably has some connection with the fact that she improved so rapidly in her adding. If we count three words for a single operation in adding, the children would be using on the average three times sixty-eight or 204 words a minute. Their reading tests show that they read on the average 210 words a minute. The fact is, however, that many children use more than three words in each adding operation. Some of them put in the word "and" or "plus" and others both "and" and "is" or "are," all of which might better be left understood in silent adding. At the rate of five words per adding operation the words used would be on the average 68x5, or 340 per minute, a higher rate than the children actually read at. The value of this

part of the analysis for the individual child is two fold. First we may show him the advantage of using three words only. Second we can hold up as a reasonable attainment for each child the rate he has already reached in reading.

We have spoken of the child's natural capacity. This is, of course, rather an undetermined quantity. In the case of adding it would seem as if besides attention, will, or ambition, susceptibility to social stimulus, etc., that general ease or difficulty in making simple associations might be a factor in the capacity to add. To arrive at this one would need to test a considerable number of different kinds of associations, and this we did not have time to do. We selected one test, the letter-square test, often called a memory test. A card with twelve letters arranged in a square was held up before the children, for twenty seconds. This was followed by a period of sixty seconds during which the children were employed in counting aloud in order to distract their attention from the letters they had just learned. When this ended the children were asked to reproduce the letters on the card. This was repeated five times and a frequency table found of the average score. A certain power to form simple associations of letters and to remember them for a short period was thus tested.

A comparison was then made as to the correlation or correspondence in standing between this test and the two tests in adding. The correlation with the first test was 48%, with the second, 57%. Here as in the other cases mentioned, practice had increased the correlation. Those with the better powers to associate letters and remember them, had on the average made greater progress during their practice period than had the others.

The analysis has concerned itself with digits up to 9. This is of course the foundation, but ordinary adding often comes in columns, in great part not reaching up to combinations of 100. This is quite a different operation mentally from simply adding a pair of digits and requires taking into account the tens as well as the units. In order to test this a list was made consisting of numbers with two figures paired with single figures, as 29, 7; 63, 8; etc. These pairs were listed in columns and the children were tested as in the other adding test. The average was only about 20 operations a minute, and after practice of nearly two weeks had arisen only to 25. The correlations of the earlier of these adding tests with the first adding test, using single digits, was 75%. After practice the

correlation had risen to 80%. It is obvious that much more practice was needed for the latter form of adding.

A consideration of these facts would seem to justify the conclusion that considerable correlation even when subjects are not closely related would indicate a general state of efficiency in the work of the school. It would show that although the pupils differed among one another, yet that each had approximately made the progress which his natural powers justified and had not been neglected in any particular subject on one hand, or driven to excessive effort on the other.

Problem Work in Arithmetic

However important the four simple operations are, they are of course only subordinate to the problem work. In order to make this real, to connect it with the children's interest and bring it into contact with their environment, we asked the children to tell us some of the things they had wondered about, and that perhaps some calculation would solve. One girl had wondered for years how many feathers were in a pillow. Another pupil had wondered how much ensilage his father's silo contained, etc. We considered how these problems could be solved. A certain number of feathers were counted and were then taken to the drug store to be weighed. The pillow full of feathers was also weighed, and from these data the number of feathers approximately obtained. The cubic capacity of a certain weight of ensilage was found, with the cubic capacity of the silo, and from these data the problem solved. In these cases the children discussed the problem in order to find what to do with it. This was done in committees with the teacher helping as the pupils required.

After the results were obtained they were presented to the class as a whole by the individuals or committees who had worked them out. The children began to work upon arithmetic as a real game which could be applied to things they came in contact with. They began to take an active attitude of investigation instead of passively receiving methods from the teacher which they were not able to apply.

Silent Reading

There is no subject in the school curriculum more important than silent reading. This is true not only for the elementary school, but for the high school and college also. In these latter institutions,

many pupils fail in their work because of their inefficiency as readers. Reading is not specifically taught in the high school; nevertheless the pupils learn to read during their high-school course. The average rate of reading for pupils at the end of the 8th grade, using easy tests is 210 words a minute. At the end of the high-school course, pupils on the average read the same text at the rate of 300 words a minute and get a proportionately larger number of ideas from what they read; while college seniors have an efficiency of 330 words a minute.

For the future citizen the capacity to read easily is more important than any other single subject he learns in school. The kind of material he reads when he leaves school is determined almost wholly by his efficiency as a reader. It is not so much the so-called sensationalism of many newspapers that leads to their wide circulation as the fact that they appeal to a class of inefficient readers. A reader who cannot read easy text faster than 100 or 150 words a minute (and there are many such even among high school graduates) will prefer newspapers where the sentences are short, the vocabulary limited, and the ideas trite and commonplace. The paper will then be largely filled with immense headlines, photographs, a kind of humorous picture writing, cheap stories and editorials in large primary grade print. If he reads the ordinary newspaper at all, such a reader will usually omit the editorials, which even for efficient high school and college students are not read faster on the average than about 250 words a minute.

To meet this situation the work done under the name of reading in the most of 8th grades is quite inadequate. In the more backward districts a book called a reader is usually provided, containing perhaps 60,000 words. If the children should or could read this at a normal rate, it ought to be finished in five or six hours. This is not the way however, the book is treated. Portions of it are too difficult to read easily and can only be stumbled over with little comprehension of the meaning. If any portion should prove to be interesting in spite of its difficulty, it is often only a small section, perhaps only a few paragraphs taken from some larger text. In the less backward districts supplementary texts are provided, but these, as well as the readers, are too often used not for the purpose of improving the children in silent reading or in getting the thought rapidly and accurately from the printed page, but for the purpose of oral reading for literary discussion. Quite often a text much

beyond the children's power to read is taken. Stanzas or paragraphs are discussed separately and reduced by the teacher to language the children understand. The pupils get the story from the teacher rather than from the text and although they may be interested in this, the process is slow and does not give them the independence of good readers.

An example will illustrate the traditional method. Before we began the demonstration work in reading, the 8th grade was occupied with the *Lady of the Lake*. The class studied or read silently one or two stanzas, after which some pupil was called on to read aloud. On one occasion when I happened to be present, a boy read:

"No thought of fear, no thought of rest
A sausaged the storm in Roderick's breast."

To the eye the right word "assuaged" looks a good deal like "a sausaged" but to the ear the difference is very marked, and the class burst out laughing, as did the reader himself. However, this was not because they really knew the word assuaged, but because they knew that sausages, at least, had nothing to do with the case. On being questioned as to the meaning, someone in the class said the word meant stopped. The teacher helped them to the word calmed, which was, of course, better. In both of these cases the minds must have worked by inquiring what is the word and the idea that is to be associated with the word and idea of storm, thus in both cases translating Sir Walter Scott's language into something more comprehensible. I asked the children at this point why Sir Walter Scott did not say calmed instead of assuaged. Some said on account of the metre, but thought the metre would have been preserved if he had said, "calmed down;" and they would have liked this better. I asked them then, "why didn't he say calmed down?" Some said, "it was because he was a very learned man and he wanted to show this off." I asked them if they liked that kind of man. They replied that they did not. Some then suggested that he put this word in so that we could learn it in school, that there were schools in those days and that the *Lady of the Lake* was probably written as a text for them. I asked them (since they thought that was the reason) whether they had now learned the word and whether they would be able to use it again; but they could think of no situation where they would ever have the courage to use this word.

I then told them the word was used in connection with thirst to mean the satisfying or partial satisfying of thirst, and asked them what could be regarded as corresponding to thirst in this case. Someone said at once that the storm referred to was the storm of anger and that this was like a thirsty beast. They began to admire Sir Walter Scott's use of the word, but still thought it unlikely they would ever use the word themselves.

If there were only a few words of this character in the text, it might be suitable enough to develop their meaning at length, but we found that there were some thirty words in every canto which the children either did not know at all or knew very imperfectly indeed. It is very important for the purpose of reading to add words to the children's vocabulary, but the words added should bear a relationship to the words already in stock both in kind and in quantity and should as much as possible consist of words capable of being used frequently by the children. To do otherwise is to miss the mental region or budding point at which the children can grow most rapidly and efficiently. It did not seem to us that the *Lady of the Lake* was the best material for this particular 8th grade, if we wished to give the kind of practice in silent reading that would be sufficiently independent and self satisfying to be continued when the pressure of the school and the constant aid of the teacher would be removed.

As a preliminary to further analysis of this problem we proceeded to test the children's silent reading capacity.

It is comparatively easy to get a test that will determine the average reading capacity of a whole class in comparison with other classes of the same or different grade, but it is a more difficult matter to determine reliably each child's separate standing. We selected reading matter from two sources. Narrative and descriptive material from *Irving's Sketch Book* and more abstract newspaper material from *Current Events*, of which we had copies every week for every child. The children were asked to read for one minute and to write out immediately afterwards a full account of what they had read. The number of words read in the text gave us the score for speed in words per second, and the number of words written in the reproduction, after repitetions and errors had been crossed out, was taken as a measure of the children's comprehension. In some preliminary tests we had found that counting the details or ideas of the reproduction gave us about the same results as counting the

number of words written (as Starch has already found). We, therefore, followed the latter method as it was simpler and occupied less time.

The correlation (or correspondence in standing) between speed and comprehension in a single test lasting one minute was not very high. For the material we used the average correlation of four such tests was found to be only 52%, ranging from 42% to 78%. The latter correlation was found in a case where the reading material (from the *Sketch Book*) was full of action, easy to grasp, and therefore easy to remember. Figure 3 shows the correlation table for this case. The correlation evidently means that on the average the child who reads quickly has passed over more ideas and therefore has more ideas to reproduce than the child who reads slowly; or to put it from the other side, the child who comprehends a given idea or detail in the text sooner than his neighbor, will pass on to the next idea or detail more quickly and thus will be reading at a more rapid rate.

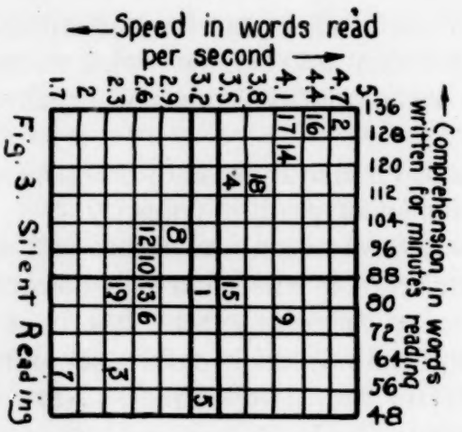

Speed in words read per second
1.7 2 2.3 2.6 2.9 3.2 3.5 3.8 4.1 4.4 4.7 5
136 128 120 112 104 96 88 80 72 64 56 48
Comprehension in words written for minutes reading

Fig. 3. Silent Reading

There are many factors however which reduce this correspondence between speed and comprehension, especially when the latter is measured by a written reproduction. It takes some ten or fifteen minutes to write out what has been read in a minute. During this time a good deal that was in the mind at the time of reading, has lapsed. If it were read again the portion might be recognized, but it is not in a position to be voluntarily recollected and written out. Sometimes whole sections will thus drop from the child's mind, and

although he knows there is something further to be recovered, he is not able to get a clue. Occasionally after he has written his paper, perhaps when he is leaving school, inhibitions are removed and the passage will suddenly flash upon him. In such cases as these the child's record for speed will stand relatively higher than his score for comprehension, and if this happens among a number of children not equally but in varying degrees, the correlation or average correspondence between speed and comprehension will be proportionately reduced.

But it is not lack of memory only which reduces the correlation. There are different degrees of comprehension even at the moment of reading. How much comprehension of a given portion is necessary in order to release the reader's attention for the next portion? With some readers, very little. I once knew an old man suffering from senile dementia who was very fond of reading and would spend hours with a book in his hand (mentally pronouncing the words, since one could often see movements of the lips and throat,) but who could give absolutely no account of what he had read even when the passage contained but two or three sentences. There must have been practically no associations of meaning aroused by the words he pronounced, and yet the practice of a well established motor-auditory habit succeeded in giving him pleasure and satisfaction—perhaps something like that of twiddling one's thumbs. This is of course an extreme; but there are a good many readers who approach it, if not for all parts of what they read, at least for certain portions. In such cases speed of reading may be rather rapid, but since there is little to reproduce the score for comprehension will not correspond; and when this happens for different pupils in varying degree, the correlation will necessarily fall.

The opposite also obtains where speed is slow and comprehension relatively large. Here there may be mechanical habits such as irregular and too frequent eye movements, or the habit of too complete pronunciation, both of which lower the speed while the mind makes use of the extra time thus forced on it in attention to details. Or, again, the comprehension may be exuberant and imaginative in the sense of revelling in details or situations of a more or less subjective character, and not strictly necessary to the understanding of the text. In such cases the reading rate is somewhat slow while the reproduction is copious if not always accurate.

The best reader falls into neither of these extremes. His attention does not dwell too long on any detail. He builds out the meaning of the text and organizes it into wholes. Meanwhile the mechanical process of moving the eye goes on with great regularity while the mental pronunciation follows behind somewhat elastically, at times being two or three words behind the eye and at other times ten or a dozen according to the difficulty of comprehension.

As will be seen from the analysis above given, the measure for speed on the average, indirectly and approximately, measures the comprehension and vice versa. The average of the two ought therefore to give a fair measure for reading capacity as a whole. In order to get numbers that are of the same dimensions and will therefore average, it is necessary to reduce the score for speed to comprehension, or the score for comprehension to speed. On the correlation table No. 9, for example, stands at 4.3 in speed and at 77 in comprehension. If we wish to translate this latter score into speed, we should take the corresponding position in speed. Since No. 9 is in the fourth column from the bottom in comprehension, we should therefore substitute the score of the fourth column from the bottom in speed. This is 2.7. The average of this with the average for speed actually obtained (4.3) is 3.5. The single score of No. 9 for reading efficiency on this particular test is therefore 3.5 words per second in terms of speed.

It will be seen from the correlation table that for this test the children write as their proof of comprehension about half as many words as they read (accurately 45%). This is for one minute's reading. Starch used half a minute's reading and found for the 8th grade a comprehension of 49 words in half a minute and a rate of four words per second. The number of words written was therefore 41% of what had been read. The total efficiency, however, is greater, but we tested the children later with Starch's test and found that their average was equal in this case to that found by Starch. It would seem therefore that reading for one-half minute gives a higher record but that this comes in from the side of speed rather from that of comprehension. As is well known the first part of a piece is better remembered than the last part. Merely to cut off the last part, however, does not seem to have the effect of giving a larger proportional reproduction. The lower proportional reproduction in the case of the half minute's reading may be due to

the shortness of the piece which would prevent the children's getting enough of the story to build out a thoroughly coherent picture or general conception. Many touches required to illuminate the understanding of the earlier part of the reading come in the latter part. The earlier part, therefore, has the advantage of this emphasis which amounts to a kind of repitition in memory and this would be one cause for the earlier portion being remembered better.

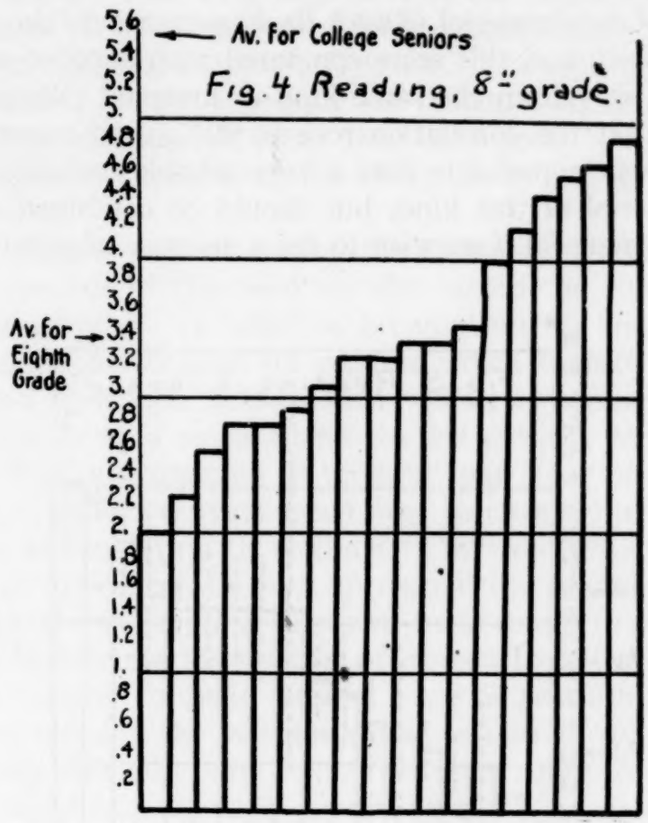
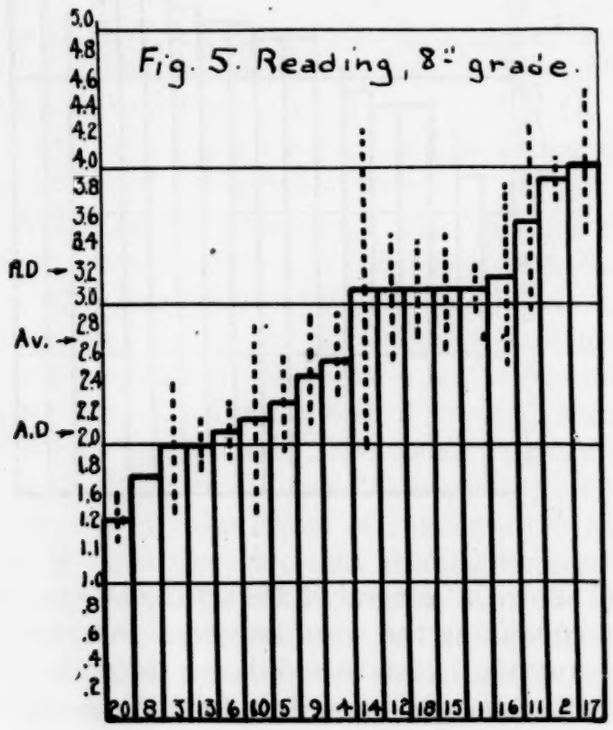


Figure 4 will show the general efficiency of each pupil in a single score made by combining the score for speed and that for comprehension. The average is 3.5 in words per second. The average reading efficiency, using the same test for college seniors is 5.5 in words per second. Another test using the same kind of material (from the *Sketch Book*) gives an average for the 8th grade of 3.6. The correlation between these two tests is 85%, thus indicating their reliability.

These tests, however, are from the same kind of material and will favor some pupils at the expense of others. This is shown by the fact that when a test from *Current Events* was correlated with one from the *Sketch Book*, we found the correlation fell to 40%—as compared with 85% with a pair from the *Sketch Book*, and 75% with a pair from *Current Events*. This means that there were some children who could read the *Sketch Book* easily as compared with *Current Events* and vice versa (allowing of course for the greater general difficulty of *Current Events*). When, however, two tests from the same kind of easy material (*Sketch Book*) were taken, the individual scores averaged, and this score compared with another score made in the same way from the same kind of material (*Sketch Book*) it was found that the correlation rose to 90%. An average of two tests from easy material is thus a very reliable individual test for reading material of this kind, but should be combined with tests from harder material if we wish to get a measure of general reading ability.



To arrive at this we took two tests from the *Sketch Book* and two from *Current Events*. Figure 5 shows the results. The individual average deviations for the four tests are shown by the vertical

dotted lines running through each score. The average for these taken together is .43 (average intra-variation). Compare this with the average deviation for the score as a whole (.62 inter-variation) and one realizes what a large extent of the total range may be covered by one pupil. No. 14, for example, in one minute's reading, may nearly be at the top of the list and in another nearly at the bottom. Great variability of this kind, however, throws an interesting sidelight on the general makeup of the pupil. This variability is characteristic of a large part of this pupil's work at his present stage of development. He is enthusiastic about some things, bored and inattentive with others. His mind is filled with a certain class of images, especially the romantic and exciting kind, while the more abstract and sober ideas are scarce. His high scores in reading naturally come from the *Sketch Book;* his low scores from *Current Events.*

A much simpler way of testing the children's reading capacity would be attained if the need for the record for comprehension could be eliminated. In order to accomplish this end the writer had a machine constructed for reading which exposed succeeding sections of a newspaper line in such widths as correspond to the widths taken in at a single glance by an efficient reader. On a newspaper line this amounts to three or four stops per line. The machine is so geared that its speed may be changed without interruption to suit any rate of reading. The slow reader may thus make two or three stops for each aperture if the machine is running slow enough, but if its rate is increased he is forced to move on and will probably decrease the number of stops. If the rate is too fast he will not be able to make sense of what he reads, but there is a point in the speed of the machine where he sees all the words, pronounces them mentally, and is able to understand their meaning The most rapid rate at which he can do this is of course his rate of reading for the material in hand. The number of lines per minute at which the machine is running may be read off on a scale and from this the number of words per second easily calculated. As a farther control the machine may be stopped at any moment and the reader asked to say what he has read. Practically this is not necessary as the reader soon complains if the machine is going too fast for him to follow. In any case there is no need for a special score in comprehension.

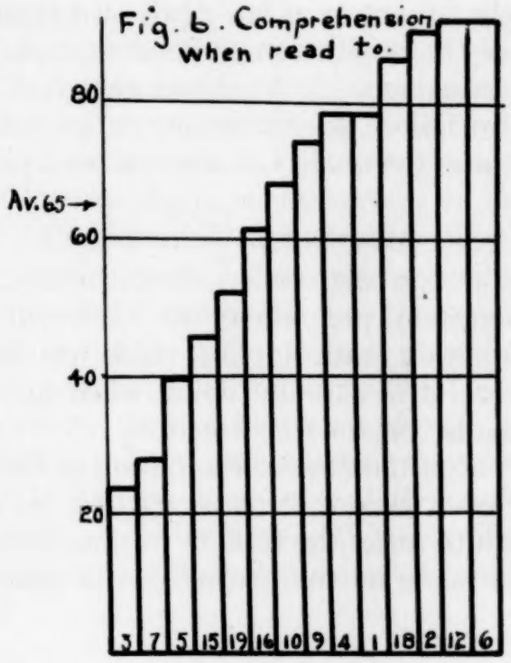
Apart from its facility in testing individuals, the reading machine is of value in improving the habits of slow readers in as far as these are due to irregular and too frequent eye movements or to complete and painstaking mental pronunciation. By gradually gearing up the machine with readers of this class, a very great increase in their rate of reading actually results, the average going up with easy texts to 4.5 words per second. At such a rate the attention must be on the alert at every moment. There can be no idling and returning to words already seen. The rate of eye movements must be regular and the tendency must be to approximate the eye movements of a good reader.

The machine is useful moreover in analyzing the difficulties of readers in as far as these are due to lack of comprehension. If the machine is running a little faster than is comfortable for the reader, phrases occur which although pronounced are not fully understood. An instance will make this plain. A pupil had read a sentence containing a statement "Our soldiers had gone abroad on a mission of patriotism." The machine was stopped at the end of this sentence and the reader asked to say what she had read. She gave the meaning of the sentence but left out the idea expressed by the words "mission of patriotism." "But there is something you have left out," I replied. "Yes," she said "there was something else I didn't get." The machine was returned to the beginning of the sentence, and she read it again at the same rate as before. But it was not until she had read it for the third time that the meaning of the phrase dawned upon her. I asked her why she had not understood it and whether she had known the meaning of the words "mission" and "patriotism." Oh, yes, mission has to do with missionaries. We send money to them in church, and patriotism has to do with Red Cross drives, the flag and love of one's country, but I didn't see how soldiers could have anything to do with missions."

This is a typical instance. The difficulty does not consist in the vocabulary as ordinarily understood. It consists in the unusual combination of words, each of which may be known but out of which the imagination must construct a meaning somewhat new to the reader. After this association is once learned it of course becomes a part of the habitual memorized equipment of the reader, and the meaning of the phrase is obvious at once. It is the breadth of associations that words have with varieties of different situations that makes the easily comprehending and rapid reader. Learning a single dictionary meaning therefore is not sufficient.

READING ALOUD

Comprehension of material which the pupil reads for himself leads us to inquire as to his comprehension when he is spoken to or when he is read to by someone else. In order to investigate this, two easy pieces were read to the class from the *Sketch Book*, each taking one minute to read. One piece was read at the rate of 143 words per minute (2.3 words per second) and the other at the rate of 157 words a minute (2.6 per second). This is of course a slow rate, but there are some in the class who read for themselves at this rate. Moreover, oral reading is slower than silent reading since the words must be fully articulated. Each of these readings was reproduced in writing by the class and the scores averaged. The results are shown in Figure 6. The average is only 65 while the average for self reading for the same sort of material is 94. This is due evidently to less ground being covered since the average rate at which the pieces were read (2.4 per second) corresponds to a reproduction of 70 words for pupils who read at this rate (see Figure 3). It is evident that the most of the class have been sacrificed to the slow rate at which the pieces were read, and they have not made up for this by remembering a greater number of details. No. 2, for example, writes only 84 words, while when she reads for herself she writes 134.



Reading aloud is a capacity which has for its principal measure of success the comprehension of what is read by those who hear. No elocutionary antics or correctness of position will make up for lack in this respect. Distinct articulation with a speed suitable for the hearers is the principal factor. In this connection it is important to notice that the rapid silent reader does not make a good oral reader without special training. In a previous investigation on normal school students (Boston) I found that there was a negative correlation of 75% between efficiency in silent reading and articulation as measured by the teacher of reading. Those who are best in silent reading, on the average, are those who do not pronounce the words fully when reading to themselves. This habit evidently goes over into oral reading. There is therefore great need for special training in this branch of the subject.

The test for comprehension during oral reading is significant not only for this exercise, but it throws a light on all comprehension of spoken discourse and therefore makes reading of the utmost significance for every subject in the school. When the teacher is speaking, some pupils are comprehending what she says much more rapidly and perfectly than others. Almost instinctively she is led to speak either at such a rate as to accommodate the slowest in the class or to use language so denuded of associations and infantile in structure as to be comprehended by her worst pupils. It is the latter alternative she usually follows, and thereby sacrifices the remainder of the class. Merely to remove a few of the worst pupils is no cure. The gradations of the pupil's comprehending capacity, as the tests show, are almost continuous. When one or two of the worst pupils are eliminated by failure in promotion or by transfer to special classes, this will raise the quality of the teacher's diction but a very few degrees.

Methods in Reading

As we pointed out in the section on arithmetic, analysis of the situation is a necessary preliminary to successful methods. The analysis of the situation in reading just made was necessary in order to show what a child is actually doing when he reads and what constitutes a reliable test for this function.

It is of course out of the question to put all of this analysis before the children, or even before the average teacher. The children, however, can learn to make the tests in reading themselves, and the more official tests made by the teacher can be placed before them.

These scores will themselves provide a considerable amount of analysis which the children can understand perfectly and which they are willing to make a basis for their practice.

After the first tests were presented to the 8th grade demonstration class, we consulted with them as to the best way of improving their reading capacity. As a result of mutual suggestions, we agreed that each child should read at home (daily) a selection from something he was interested in and was perhaps already reading, and give an account of this orally to his classmates. He was asked to time his reading and report this also. How much can be told next day of something read once and taking ten minutes to read? In order to give a proper and intimate audience, and in order that each child would have an opportunity to give an account every day, the class was divided into groups containing four or five pupils each. The pupil after he had given his account, asked the others what images or ideas they had gotten and emphasis was placed on the differences of these individual pupils. The hearers then asked questions as they wished and often brought out parts of the account which had not been given at first, a practice which tended to make this pupil's account better the next time. Criticisms were also offered as to the manner of speech. Too many hesitations—too many "and's" —couldn't hear him—didn't open his mouth enough—bad grammar— were some of the more frequent criticisms made. The group then decided which account had proved the most interesting and this was recommended to be given to the class as a whole.

After the pupils had made some progress, it was suggested that they bring the book or magazine from which they had made their selection, to school, and where suitable, read the actual words of the text, putting these into the body of their spoken reproduction.

Reading from *Current Events* was taken up by having pupils read silently for about a minute and then single pupils would be selected to give an oral reproduction. Time was not allowed for re-reading or for slower pupils to read as much as the faster ones.

Besides this, many pupils began to read somewhat harder material, such as was contained in the editorials of the daily papers. They noted any phrase with which they were not familiar and which delayed their reading. They wrote these phrases down, or sometimes clipped the line containing them out of the newspaper and thus made lists of them in a note book given them for this purpose. They brought these phrases into their groups and if members of the group

did not know them they had the privilege of adding them to their collection. The following by No. 17 is a small part of such a collection.

radical theorist
grovelling terms
peace without annexations or indemnities
vehement, tumultuous emotions
clarify the illusions
imperceptibly sweeten life
possessing the originality
unsightly and unnecessary projections and appendages
muddle-headed ignoramus
congenial companion
by their consituents
garrulous exultant throng

Another from a less efficient reader, begins as follows:

practical conclusions
general principles
Prussian oppression
in substance
radical theorists
enigmas of the war
future exigencies
commercial aspirations
congenial companions
sentiment of the country
it is incomprehensible
ingredients of clay
every one was surfeited

There is a striking difference in the character of these lists. Probably no phrase in the last list would have given any trouble to the collector of the first list, and the most of the phrases in the first, would have been unsuitable for learning by the second pupil until she had mastered commoner phrases such as she actually collected. Each pupil is thus learning individually at his own level, but what one learns is also communicated to the others and shared by them in as far as their interest and ability warrant. They are free to put phrases collected by others into their list or to leave them out, as they see fit. They were recommended not to put phrases into their lists which they did not feel they could find some opportunity to use either in conversation or in their compositions. They frequently reported some of their attempts to use their newly discov-

ered phrases at home. One boy said to his father, "What do you think of the proposed ratification of the prohibition bill?" His father said, "What's that?" On its being repeated, the father said, "You may as well talk German at once, and have done with it." Since the boy told his story with good humor, it is likely that the reaction of his father rather stimulated him than the reverse. He might have felt that he had scored one on his father, or he might have seen a lurking admiration in his father's seemingly unsympathetic answer. A girl reported that she had asked her mother to "clarify her ideas" on some matter. The family twitted her all the evening, asking her if she could clarify their ideas on this subject and on that, but it is likely that the process had helped rather than hindered.

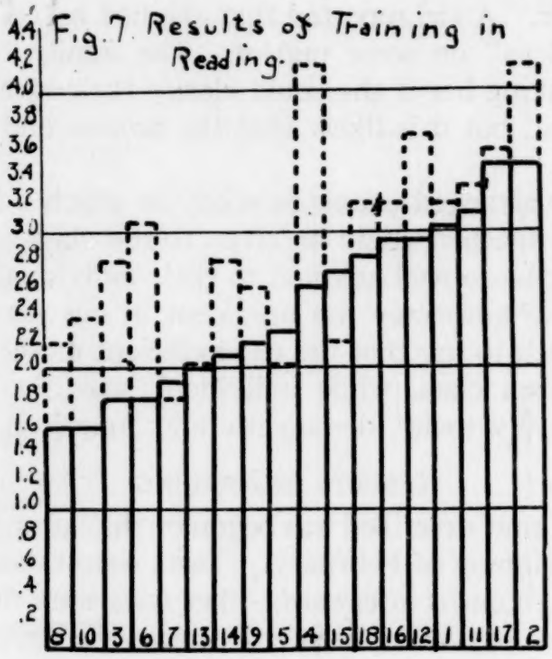
A previously arranged course of study in which a list of phrases selected by the teacher would be given to the class and learned by all, could never be so well adapted to their individual needs. This is the method of autocracy too prevalent in our schools. I think it is not too much to say that the more efficient method followed by the demonstration class, while individual, was also co-operative, and was thus truly social, democratic and American.

Results in Reading

Work of the kind described was begun in the latter part of January and the beginning of February. Tests were taken at this time and about three months afterwards other tests were made as nearly like the first as possible. Figure 7 shows the difference. Each score represents the average of two tests one minute in length from *Current Events*. The average of the first test (score combining speed and comprehension) is 2.4; that of the second, three words per second. The correlation of the two tests is 85% showing the class had progressed pretty evenly. The average deviations (inter-variation) are 5.4 and 5.6 respectively. These overlap somewhat but not enough to weaken very much the independent status of each average.

The progress is not so striking in reference to the total achievement as that already shown in the case of arithmetic, but the reason for this is that the total volume of associations to be learned in the case of reading are hundreds of times more numerous than in the case of adding. As we found in a vocabulary test, the average reading vocabulary was about 21,000 words in a test made from a

large dictionary, in which the average of college seniors was 42,000 words. When one adds to this vocabulary the many associates with each word, we have a number of mental operations compared with which the 81 associations to be learned in adding represent a very small task indeed. The proportion therefore that the children will learn in three months' time cannot be expected to be very large. Perhaps at the rate of a little over 2000 words per year.



As we have already seen, vocabulary is only one factor in reading. This is shown by the fact that when two tests in vocabulary were averaged, of which the correlation was 71%, and this score correlated with the score in reading, made also from an average of two tests, (correlation 85%) the correlation between vocabulary and reading amounted to only 52%. This shows that while vocabulary is distinctly a factor, eye-movements, habits of mental pronunciation, familiarity with phrases, interest in the content, imaginative construction must also play a large part in reading efficiency.

As the figures show, however, the progress is noticeable, amounting to a difference of .6 words per second or 36 words a minute. It is as much, in fact, as Starch found on the basis of a large number of schools throughout the country for the average growth of a whole

year between the 7th and 8th grades. That this progress was made during a practice period of three months must be attributed to some superiority in the process at work.

Composition, Spelling, Writing, and Grammar

A good part of the work the children did under the head of reading can also be classified under composition. The extension of the children's vocabulary and phrases and their practice of making reproductions both written and oral, certainly belong here, as well as under reading. But this kind of work is largely passive and is concerned with the ideas of other people. An opportunity for something more active and as creative as possible is also required.

Composition as it is often conducted in schools is not regarded as primarily the art of written communication. Children are not usually asked to write to anyone in particular unless perhaps the teacher; and in this case when the children are as old as those in the 8th grade, they frequently see through the illusion and realize that they are actually writing for the waste paper basket. When a business man writes a letter, he has some object of a social character in view. He is writing to someone and he is expecting to cause some change, however small, in that person's mind. In harmony with this purpose he selects his ideas and chooses his language. Lack of clearness, of force, deficiencies of spelling and grammar, are avoided because of their effect on the mind of his reader. He is, moreover, able to gauge the success of his effort by the response which he obtains. All of these motives for efficient and interesting writing are usually absent in the average school.

After testing the children by the Harvard-Newton scale, and finding that no composition was better than the fourth class of this scale (which runs from 1 to 6) we suggested to the children that they use the same method they had been using in reading, that is, the children wrote compositions on subjects of their own choosing which were then divided into sets of four or five and handed to groups or committees of the same number. Each member of a committee then read each paper handed to his group (none of which were his own) and arranged the various papers in order of merit, according to his judgment. When every paper had been read by each child, the committee began to discuss the merits and demerits of the composition and to explain why each had ranked the set of papers as he had. If in this discussion any child saw good reason to change

his ranking of the papers, he was at liberty to do so. The average standing of the different compositions was then taken by adding together the rank given by each individual reader and dividing by the number. This was reported to the teacher and the class and the best composition recommended to be read by the writer himself to the class as a whole. While the discussion as to the comparative merits of the papers was going on, the teacher passed from group to group, taking part as she saw fit. She was frequently appealed to on various points of structure, grammar, usage, etc., and took the opportunity thus afforded to give the children the instruction desired. The different sets of papers were then taken to the College and a group of five or six seniors read and ranked them in the same way the children had done, but without knowing previously what the children's ranking had been. On no occasion was the average ranking on the part of the students different from that already found by the pupils of the 8th grade.

We asked the children whether they thought any of their parents or other relatives would care to read some of these sets of papers and rank them as the college students had done. The children thought a letter from the teacher to the parents would help matters. It was suggested that the class compose a suitable letter, which, after passing everybody's criticism, would be signed by the teacher and sent with a set of compositions to the various homes. This was done, not so much to get a still more objective standardization of the papers, but to interest the parents in the work of the school.

In order to test the power of 8th grade children to judge compositions, not their own but written by other 8th grade children, we had copies made of a descriptive set in the Harvard-Newton scale. The children were asked to grade these as they had been doing already with the compositions of their own class. The judgments of one group are shown in Table I. The other groups gave similar results. The order given to these papers by the eighty 8th grade teachers, who judged them first (although their variation was large) is the same as the order found by the children, except that the "Lake at Sunrise," and "The Mansion" are in reverse order. These compositions are, however, practically paired by the children. My experience with adult judges, moreover, shows that these two compositions are frequently interchanged and are actually harder to discriminate than any of the others. In a previous investigation made on the 8th grade of the Everett School in Boston, I found

further evidence of the capacity of the 8th grade children to judge these compositions. Of 37 pupils in this grade, the average judgment was for "The Storm in the Fishing Village," 1.84; "The Grandmother," 2.03; "A Mansion," 3.13; "The Lake at Sunrise," 4.05; "The Lighthouse," 4.50; "A Scene on the Prairie," 5.35.

TABLE 1. PUPIL'S JUDGMENTS OF COMPOSITIONS

Correct order (Teachers)	Pupils judging	3	5	2	20	9	19	Av..
5	The Lighthouse	4	3	5	5	5	5=27	4.5
2	The Grandmother	2	2	3	1	2	3=13	2.2
3	A Mansion	3	5	4	2	3	4=21	3.5
1	Storm in the Fishing Village	1	1	1	4	1	1= 9	1.5
6	A Scene on the Prairie	6	6	6	6	6	6=36	6.0
4	The Lake at Sunrise	5	4	2	3	4	2=20	3.3

It would seem from these facts that a random group of five or six 8th grade children would in the most of cases arrive at the same, or nearly the same, average judgment as the teachers of this grade.

When the next compositions were judged, the sets of papers and groups were differently arranged, so as to give each pupil a chance of being compared with other pupils in the class. A pupil who had been judged by a committee in one case as first, might be judged, when compared with other writers as second or third, even when his paper was equally good. In estimating the pupil's real capacity these conditions were naturally held in mind, both by the writer and his judges.

There is something to be said in favor of judging the pupil in comparison with a small number of his classmates rather than with all, even if this latter could be done successfully either by the teacher or the pupils; because although the pupil receives some honor or the reverse, there is no reason for him to be either puffed up or discouraged. It will only be after a large number of such tests, that he and his comrades may estimate finally his general ability as a writer. Meanwhile there is plenty of opportunity for him to try various ways by which he may please and impress his readers.

Work of this kind combined very well with the work in reading. I was present once when the best compositions from different groups were being read to the class as a whole. Approval or the reverse was expressed by the members of the class. "I like your saying stately maple, you might have said, "stately maple tree," or "Did you get that phrase 'the mellow light of the moon,' recently for your collection?" "Yes," answered the writer.

As has been said, at the beginning of this work we could judge no composition higher than the fourth position on the Harvard-Newton scale. After four months' work, four writers at least had moved up to the third position, while only two or three were still left behind in the fifth. Lack of space prevents us giving samples of these compositions.

While this work was going on, we tested the spelling, the penmanship and the formal grammar. In the latter subject the children stood high, although there was no particular evidence of it either in their speech or conversation. In spelling (Ayres' scale) they were found to be about a year behind their grade, and in penmanship, about on the average. We did nothing special to improve these subjects, except in as far as they entered into the composition. The teacher meanwhile conducted recitations in these branches as formerly.

Neither had we time to make anything more than a beginning in history and geography, subjects which promise large returns, both from the standpoint of increased efficiency and co-operative democratic methods. The manual training, drawing and music were taught by special teachers and were unrelated to the rest of the school work. Nothing was done in these subjects, either, towards increasing their efficiency. We look forward to the work of another year to continue what we are now doing and to add some other subjects to our field of demonstration.

The work of the year was made possible and enjoyable by the loyal and intelligent co-operation of Miss Marion Chase, the teacher in charge, and every courtesy was extended by Mr. Whittemore, the superintendent. It is expected that this demonstration class will be made a station for observation first for the teachers of the village and secondly, although indirectly, for other teachers throughthe country.

THE EFFECT OF EXEMPTING PUPILS PROFICIENT IN HANDWRITING

CYRUS D. MEAD

University of California

The pupils of an ordinary fifth or sixth grade class will write with varying degrees of legibility. There will probably be some pupils as low as quality 6 or 7 and some as high as quality 13 or 14 by the Thorndike Scale, this being nearly the complete range on the Ayres Scale. It has been found that, roughly, one third of the pupils of any one grade, from the first to the eighth inclusive, will reach or exceed the median quality of the next higher grade; and that one fourth of the pupils will reach or exceed the median quality of the second grade above it. For any other subject than writing, this condition might become the basis for a differing gradation. It is much less a problem in writing in that class instruction and practice in this subject may include a wider range of abilities without impairing the efficiency of such instruction. But why should individual pupils who have acquired the writing abilities of so called "standards" above their grade and show a reasonable speed, be required to take formal instruction in writing so long as they maintain such a quality? The very suggestion however, of exempting such pupils from "regular writing periods" will, almost without exception, be met with indignation by a supervisor or teacher of writing. It is as if some sacred custom had been violated beyond repair. If you protest that you have no pre-conceived notions regarding the matter but that you wish to get experimental evidence upon which to base some notions as to the effect, you are apt to be told that handwriting is too important an accomplishment to be experimented with. Yet such an attitude of mind prevailed in science from Aristotle to Galileo, and history records many of the intervening centuries as the "Dark Age."

To find whether or not the posting of a handwriting scale in a class-room was any incentive to improvement in writing and whether

---

[1]With the cooperation of class teachers from day to day, the following members of my Educational Measurements Seminar in the College for Teachers of the University of Cincinnati conducted the preliminary and final, and also the intervening monthly progress tests:

C. M Howe in the Fairview School; John W. Snyder in the Madisonville School; and John F. Arundel in the Lincoln School, Cincinnati.

or not exempting especially proficient pupils was detrimental to them, and to class progress in writing in general, an experiment was carried out over the greater part of the school year of 1917-18 with three fifth and two sixth grades, in all, 203 pupils.[1] The problem and plan stated briefly were as follows:

A CONTROLLED TEST IN HANDWRITING

Problem: 1—Does the quality of handwriting in a class suffer by exempting especially proficient pupils from formal writing periods?

2—Does the posting of a handwriting scale effect improvement in writing?

Plan Three classes each of fourth, fifth, and sixth grade pupils selected and divided into three groups as follows:

Group I—One fourth, one fifth, one sixth grade.

1. Sixty minutes weekly for formal writing exercises.
2. Gettysburg Scale posted and explained.
3. Pupils encouraged in using same for comparison of work from time to time.
4. Monthly tests for scaling.
5. Those pupils who reach speed and quality of one grade above exempted from practice at formal writing periods, this dependent however, on their maintaining same in each monthly test.

Group II—One fourth, one fifth, one sixth grade.

1. Same
2. Same
3. Same
4. Same

Group III—One fourth, one fifth, one sixth grade.

1. Same

(A preliminary and final test)

The original plan as seen above contemplated three groups of classes, a fourth, fifth and sixth grade in each of the three groups and three different schools. Due to various difficulties, the final data could be completed only for the three fifth and two sixth grades. These classes were of the regular school organization and from the nature of things not necessarily of equal abilities within their re-

spective grades. They also had different teachers. These are qualifying factors for which the reader must allow. It will be seen however from the scores below that Groups I and II, that is, the fifth grades of these groups, in the initial tests, were of practically the same average and median abilities in quality of writing. It is in these groups that the chief interest centered. Groups I and II were "German" sections. Some time during the day was given over to instruction in German. (This has since been done away with). These groups had the full benefit of the posting and featuring of the Ayres Gettysburg Handwriting Scale and of monthly tests therewith. The chief factor of difference however was that the pupils of Group I had the added incentive of exemption of pupils each month from all formal writing periods when such pupils showed sixth grade ability, and held it, in both speed and quality.[2] This gave them the three twenty minute periods weekly for other more needed work.

Group III (the fifth grade) was a regular "English" class with no instruction in German. It was probably a slightly inferior group in that it included some failures from the classes taking German. The average and median initial scores in writing show them to be about one point lower on the Ayres Scale, although more rapid writers. This group had no knowledge and made no use of the scale in the class-room. It was given only the initial November and final May tests. Condition one, however, was the same, *i. e.* three twenty minute formal writing periods weekly. The purposes of the controlled tests were, again, to obtain some objective data for altentative answer to two questions: First, did the incentive used of posting and explaining a handwriting scale in the school room prove to be an effective agent in improving handwriting as shown by comparison of Groups I and II with Group III; second, did the described exemption from all class practice of proficient pupils lower or raise the standards of their groups, as shown by comparing Groups I and II with each other? The results with the fifth grade classes will be taken up first. When referring to groups, this grade only is meant. For purposes of illustration, the records of one class are given in Table I. The intervening monthly scores are omitted but the leaded lines mean the months over which the pupil was required to take formal instruction and practice in writing. A space means that the pupil was exempt. All tests were given during the first week of each month.

[2]By sixth grade ability is meant the average number of letters written per minute and the average quality of writing "commonly found in grades".

The following averages are taken from the Ayres Gettysburg Scale:

Grade	Speed	Quality
2	31	38
3	44	42
4	55	46
5	64	50
6	71	54
7	76	58
8	79	62

TABLE I

Handwriting Records—1917-18—Miss Dora K's Room—Fairview School (Black line means pupil took regular writing instruction; space means pupil exempt)

Group I,	Nov.–1917		Dec.	Jan.	Feb.	Mar.	April	May 1918	
Pupils	Sp	Qual						Sp	Qual
C. L.	80	72.5						86	82.5
H. E.	72	70						78	60
J. R.	76	55						77	60
L. M.	65	60						70	62.5
D. M.	63	67.5						86	57.5
D. S.	78	52.5						88	62.5
G. W.	81	50						78	52.5
H. A.	72	50						75	60
M. T.	75	50						69	47.5
C. V.	91	50						81	62.5
M. B.	60	57.5						77	60
R. W.	62	55						75	62.5
T. B.	53	50						63	45
E. F.	53	57.5						57	52.5
M. H.	54	60						88	62.5
C. M.	52	50						78	60
M. G.	47	55						74	65

C. H.	41	62.5	67	60
L. S.	52	52.5	91	52.5
E. H.	60	52.5	88	62.5
C. N.	53	57.5	76	62.5
M. V.	72	45	72	65
J. R.	76	40	57	52.5
F. M.	74	35	76	42.5
H. D.	63	47.5	70	60
W. S.	65	42.5	88	45
P. G.	67	40	77	50
G. C.	65	40	70	50
A. S.	63	37.5	70	45
E. K.	53	40	73	55
A. D.	57	37.5	74	50
R. W.	46	45	73	60
R. K.	54	37.5	54	37.5
H. C. K.	57	42.5	75	40
R. B.	60	40	66	42.5
J. P.	60	35	69	45
R. H.	66	40	62	52.5
M. N.	63	40	105	52.5
G. K.	56	35	87	47.5
S. W.	54	40	64	52.5
H. M.	56	45	76	47.5
G. M.	52	42.5		
J. G.	74	37.5		

Tables II, III and IV assemble in frequency table form the initial and final scores of the three classes in both speed, or letters written per minute, and quality as measured by the Ayres Scale. Below each table for purposes of comparison are recorded the common measures of central tendency and of variability. Table V summarizes these measures and records the changes effected from November to May in each of the three groups.

TABLE II

Group I. Fifth Grade—Speed and Quality Scores—Initial and Final Tests

Score in Letters Per Minute-Speed	Number of Pupils		Quality of Writing.	Number of Pupils	
	Nov. 1917:	May 1918		Nov. 1917:	May 1918
			25–29		
30–34			30–34		
35–39	1		35–39	7	
40–44	1		40–44	10	3
45–49	2		45–49	3	7
50–54	8		50–54	9	11
55–59	4	2	55–59	6	2
60–64	9	3	60–64	3	15
65–69	5	4	65–69	1	2
70–74	4	9	70–74	2	
75–79	4	13	75–79		
80–84	2	1	80–84		1
85–89		7	85–89		
90–94	1	1	90–94		
95–99					
100–104					
105–109		1			
Number	41	41	Number	41	41
Guessed Average	62.5	77.5	Guessed Average	52.5	52.5
True Average	62.95	76.3	True Average	49.94	56.28
Median	62.5	75.96	Median	50.28	54.80
Average Dev.	9.0	7.1	Average Dev.	8.2	6.95
Quartile Dev.	8.5	4.6	Quartile Dev.	7.4	6.24

	Growth	Speed	Quality
	Absolute	13.46	4.52
	Per Cent	21.5	9.0

From a careful study of Table V we may tentatively answer our two questions. The use of the scale did not produce as decided an advantage in Groups I and II over Group III as one might expect, though the increase in speed is enough greater in these

groups to more than balance larger growth in quality in Group III. And it is noteworthy that in every instance the variability of the Groups I and II is reduced more than that of Group III. Perhaps a fairer showing for the use of the scale may be obtained in another way. The fact that Group III started at an abnormally low level of quality, obscures the comparative effect, since the growth was so much easier at that plane. Let us now compare the groups as to increase in the number of pupils up to Ayres fifth grade standards for both speed and quality in November and May. In Group I the number showing fifth grade ability in both speed and quality increased from 9 in November to 27 in May, a net growth of 18 in a class of 41; for Group II there was an increase from 6 to 32, or a net growth of 26 in a class of 44, while in Group III the increase was from 2 to 10, or 8 out of a class of 37. As this is the big task

TABLE III

GROUP II. Fifth Grade—Speed and Quality Scores—Initial and Final Tests

Score in Letters Per Minute—Speed	Number of Pupils		Quality of Writing	Number of Pupils	
	Nov. 1917; May 1918			Nov. 1917; May 1918	
			25–29		1
30–34			30–34	1	
35–39			35–39	1	
40–44			40–44	9	
45–49	2		45–49	12	9
50–54	9		50–54	9	15
55–59	1	1	55–59	9	7
60–64	7	2	60–64	1	8
65–69	7	9	65–69	1	3
70–74		5	70–74	1	1
75–79	5	6	75–79		
80–84	9	10	80–84		
85–89	2	5	85–89		
90–94	1	3			
95–99					
100–104	1	3			
Number	44	44	Number	44	44
Guessed Average	67.5	77.5	Guessed Average	47.5	52.5
True Average	68.98	78.75	True Average	50.34	55.00
Median	67.14	79.17	Median	49.57	54.00
Average Dev.	11.48	8.75	Average Dev.	6.00	5.68
Quartile Dev.	13.06	7.78	Quartile Dev.	5.28	5.15

Growth	Speed	Quality
Absolute	12.03	4.43
Per Cent	18.0	8.9

TABLE IV

Group III. Fifth Grade—Speed and Quality Scores—Initial and Final Tests

Score in Letters Per Minute—Speed	Number of Pupils		Quality of Writing	Number of Pupils	
	Nov. 1917; May 1918			Nov. 1917; May 1918	
			25-29	3	
30-34			30-34	6	
35-39			35-39	13	6
40-44			40-44	7	11
45-49			45-49	6	10
50-54	2		50-54	1	6
55-59			55-59	1	4
60-64	4	1	60-64		
65-69	2	1	65-69		
70-74	7	6	70-74		
75-79	5	3			
80-84	8	3			
85-89		6			
90-94	3	7			
95-99		4			
100-104	2	2			
105-109	3				
110-114		3			
115-119	1	1			
Number	37	37	Number	37	37
Guessed Average	77.5	87.5	Guessed Average	37.5	47.5
True Average	80.2	88.31	True Average	39.39	46.28
Median	78.5	88.75	Median	38.65	45.75
Average Dev.	11.62	10.54	Average Dev.	5.13	5.0
Quartile Dev.	6.97	9.43	Quartile Dev.	4.51	4.58

Growth	Speed	Quality
Absolute	10.25	7.1
Per Cent	13.1	18.4

TABLE V

Changes Effected From November to May

	Speed		Quality		Variability in Speed		Variability in Quality	
	Average	Median	Average	Median	A. D.	Q.	A. D.	Q.
Group I	62.95 to 76.3 +13.35	62.5 to 75.96 +13.46	49.94 to 56.28 +6.34	50.28 to 54.80 +4.52	9.0 to 7.1 -1.9	8.5 to 4.6 -3.9	8.2 to 6.95 -1.25	7.4 to 6.24 -1.16
Group II	68.98 to 78.75 +9.77	67.14 to 79.17 +12.03	50.34 to 55.0 +4.66	49.57 to 54.0 +4.43	11.48 to 8.75 -2.73	13.06 to 7.78 -5.28	6.0 to 5.68 -0.32	5.28 to 5.15 -0.13
Group III	80.2 to 88.31 +8.11	78.5 to 88.75 +10.25	39.39 to 46.28 +6.89	38.65 to 45.75 +7.10	11.62 to 10.54 -1.08	6.97 to 9.43 +2.46	5.13 to 5.0 -0.13	4.51 to 4.58 +0.07

of fifth grade writing we may answer our first question by stating that the use of the Ayres Scale did add materially to the bringing of a large percentage of the class up to fifth grade standards, and in addition, decreased the variability of Groups I and II.

As to the effect of excusing from formal instruction proficient pupils showing sixth grade ability, comparison of Groups I and II shows that far from retarding grade progress, it increased it in both speed and quality. If in addition to this favorable showing we examine the economy of time for these proficient pupils, the case seems clearer.

Number of Pupils Excused Each Month in Group I, and the Average.

Month	Nov.	Dec.	Jan.	Mar.	Apr.	May	Aver.
Number	3	9	2	7	9	17	6.8

As this saved for each pupil one hour per week, four hours per school month, or twenty-eight hours for the seven months, the time economy was 6.8x28 or 190.4 pupil hours. May we then say that the evidence from these fifth grade tests is in favor of making use of handwriting scales in the class room and of exempting proficient pupils from needless writing practice, thus effecting, with no loss, a saving of time for other work?

Results with Sixth Grade Classes

The work of the two sixth grades is somewhat surprising in that we find them both losing in quality of writing after a year's instruction and practice. The writer feels, however, that this is due somewhat to an abnormal effort in the final May tests to "speed up," with a consequent decrease in quality, as is shown by the median results. Again, it was felt that the elements throughout the year with these groups were not as carefully controlled as with the fifth grade group. The frequency tables for the two sixth grade groups are omitted, but Table VI summarizes the measures of central tendency and variability in which the changes effected from November to May are shown.

The results in Table VI allow us mainly a choice between two evils, which did the poorer? What "works" with one class does not of necessity work with another. The interest, sympathy and enthusiasm of teachers differ and this is apt to be reflected in the groups of pupils. It is of interest however to call attention to two facts, one, that the two classes were somewhat widely different in quality of writing, Group II being of about third grade ability, on the average, and Group I of seventh or eighth grade ability, and at the close of the year, poor as their showing was, this latter group

was up to or beyond the average of sixth graders in general. And this was true after the second fact to be noticed, that is, that 9 pupils on an average were exempt each month from 45 minutes weekly writing, their time being open for other study.

TABLE VI

Changes Effected From November to May

	Speed		Quality		Variability in Speed		Variability in Quality	
	Average	Median	Average	Median	A. D.	Q.	A. D.	Q.
Group I	72.9	76.5	66.5	64.8	14.0	11.5	8.0	6.2
	to	to	to	to	to	to	to	to
	91.3	94.2	56.1	55.2	14.0	11.1	9.0	9.1
	+18.4	+17.7	-10.4	-9.6	0.0	-.4	+1.0	+2.9
Group II	72.3	71.4	43.9	42.0	10.3	8.5	12.4	10.5
	to	to	to	to	to	to	to	to
	81.9	81.4	39.4	38.3	15.4	16.9	8.3	7.1
	+9.6	+10.0	-4.5	-3.7	+5.1	+8.4	-4.1	-3.4

Number of Pupils Exempt Each Month in Group I, and the Average.

Month	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	Av.
Number	14	4	3	7	13	11	11	9

Further than this 19 pupils in Group I were up to sixth grade standards in November and this increased to 22 in May. In Group II, 2 pupils were up to sixth grade standards in November and this decreased to none in May.

While it is recalled that the results with the three fifth grades, where the conditions were somewhat more carefully controlled, were more positive, still it is far from the purpose of those conducting the tests to "prove" anything. The question is simply raised: Of what good is it to require pupils with the speed and quality in writing of grades ahead to take formal practice in writing with the remainder of the class, if they maintain such, and if the time might be employed with other studies in which they may be deficient? We may all have our opinions. These, however, may be set up in a "balance of powers." Actual data obtained under controlled conditions will allow us to arrive at a more positive, if not a more peaceful, conclusion.

COMMUNICATIONS AND DISCUSSIONS

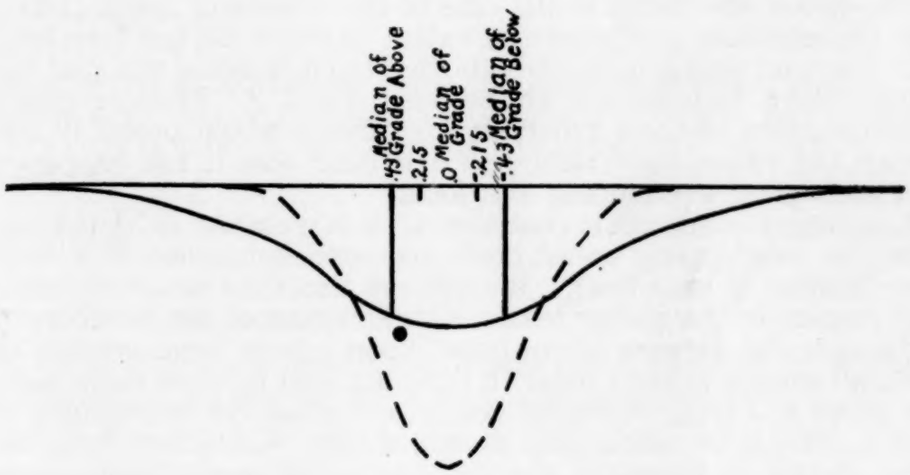
THE MEASUREMENT OF OVERLAPPING

All who have given standardized tests or have used scales for the measurement of accomplishment of children in different grades have found great variability in each grade and extensive overlapping of grade upon grade. The existence of extreme overlapping in reading ability is reported by Daniel Starch[1] who writes "it is found that in speed and comprehension 31.8 per cent of the pupils of any grade reach or exceed the median of the next grade above," and by F. J. Kelly, who reports[2] that, 'The reading ability possessed by the median child in any one grade is superior to the ability possessed by at least one third of the children in the class above him and is no better than the reading ability possessed by at least a third in the class below him."

Assuming a normal distribution of scores in reading ability and calling the standard deviation of such scores Σ, this statement would imply, as may be determined by reference to a table of the frequencies in a normal distribution, that one grade median is $.43\Sigma$ distant from the next. The situation is pictured in the accompaning full-line graph.

[1] DANIEL STARCH. *Educational Measurements*, p. 42.

[2] *The Kansas Silent Reading Tests*, J. Ed. Psych. v. 7 no. 2, 1916.



Let us assume that the score in the reading test is a valid criterion of ability, and determine the per cent of pupils who would be better classified if they were in the grade above or the grade below. Any pupil making a score greater that that which corresponds to .215Σ is closer to the median accomplishment of the grade above than to that of the grade in which he is located, and therefore should be transferred to the grade above. The percentage of cases in a normal distribution lying above .215Σ is 41.5, and similarly 41.5 lie below—.215Σ. Accordingly, by this criterion, 17 per cent are properly, and 33 per cent improperly classified. As, thruout a number of grades, reading ability is presumably the most important single determinant in classification, so extreme a finding as this may well be questioned and carefully considered before finally accepted.

We may approach this matter from another point of view. If we give a five minute test in solving puzzles we are testing ability to react to novel situations and may call our test a reasoning ability test. From data at hand the writer knows that enormous overlapping would be found with such a test. The reason is here quite obvious. The puzzle ability of an individual may be quite identical with his reasoning ability, but it cannot be determined in five minutes. The probable error of a five minute determination of puzzle ability age is very likely three or four years. The great overlapping in this case is seen to be a consequence of the unreliability of the test, and not the trivial nature of the function tested or of the real overlapping of talent from grade to grade.

To a degree the same situation maintains in the case of every measure of overlapping, unless it be in dealing with such measures as height or weight which can be "exactly" measured, and the unreliability of measures must be known before significant estimates of overlapping can be made.

The writer has found in the case of two classes of about thirty, that the reliability coefficient (the extent to which the test correlates with a second similar test) of both the Starch Reading test and the Kansas Silent Reading test was between .2 and .3. More extensive determinations of these reliability coefficients would probably not change the values appreciably, so it is clear how it has happened that such great overlapping was found.

Knowing the reliability coefficient, it is possible to determine what the overlapping would be if measured by means of a very large number of such tests. We will call this the true overlapping with respect to the ability tested. If the scores of the members of a class are the average scores based upon a very large number of tests, all similar to each other in difficulty and in what each tests, the scores will truly represent ability and whatever overlapping is then found will be overlapping in true ability. Consider the dotted curve shown to represent the distribution of scores based upon averages of a large number of tests. If the grades received in the first test are designated by x^2 (x^3 in the second test, x^1 in the third,

x n in the n'th), where x^1 is a deviation from the group mean, then the full line curve represents the distribution of grades x^1 received by the member of the class (consider N to be the number of individuals), and the dotted curve represents the distribution of the N average grades, each of which is of this type, $\frac{x_1 + x_2 \cdots + x_1}{n}$. We wish to know the spread or dispersion of these latter measures and will therefore calculate their standard deviation:

$$\frac{\sigma^2 x_1 + x_2 + \cdots xn}{n} = \frac{\Sigma (x_1 + x_2 + \cdots x_1)^2 n}{n}$$ in which Σ indicated a summationof N terms.

$$\frac{\Sigma x \Sigma 1^2 + \Sigma x_2{}^7 + \cdots \Sigma x_1{}_7 + 2\Sigma x_1 x_1 + 2\Sigma_1 x_2 + \cdots 2\Sigma x_1 x_2 + 2\Sigma x_1 x_2 + \cdots}{n_7 N}$$

Since according to hypothesis the tests which are averaged are all similar they all have the same standard deviation and correlate with each other to the same amount; that is to say

$\Sigma \sigma x_{12} = \Sigma x_2{}^7 = \cdots = N \sigma 1^2$, according to the usual definition of standard deviation, and

$\Sigma_1 x_2 = \Sigma x_1 x_3 = \cdots = N\, r\, \sigma_1 \sigma^2$, according to the usual definition of a product-moment coefficient of correlation.

In this last equation r is the correlation between one text and a second similar one. In the equation given there are n terms such as $\Sigma x_2{}^1$, and $\frac{n(n-1)}{2}$ terms such as $\Sigma 2 x_1 x_2$, so that we have

$$\frac{\sigma^2}{x_1 + x_2 + \cdots xn} = \frac{n N \sigma_1{}^2 + n(n-1) N r \sigma,{}^2}{n^2 N} = \sigma_1{}^2 (n + r - n^{1} r)$$

As n becomes indefinitely large the expression in paranthesis becomes r, so that representing the standard deviation based upon a very large number of tests, *i. e.* the true standard deviation, by σt we have:

$\sigma t = \sigma_1 \sqrt{r}$ (Formula giving the standard deviation of the average of an indefinitely large number of similar tests, knowing the standard deviation and coefficient of reliability of the single tests)

In the case of the Kansas Silent Reading tests, taking the reliability coefficient = .25 and the standard deviation $\sigma 1$, we have

$$\sigma t = \sigma_1 \sqrt{.25}, \text{ or } \sigma t = .5\, \sigma_1$$

In the original distribution of scores it was necessary to go out from the mean a distance of $.215\sigma$ to reach a score half way between the median scores of two grades. It will accordingly be necessary in

terms of the distribution of true scores to go out $\frac{.215}{.5}(=.43)$ standard deviations to reach the point midway between grades and to go out .86 0 to reach a neighboring grade median. Therefore according to the overlapping reported by F. J. Kelly there is 33 per cent, instead of 17 per cent, of correct classification and, referring to the frequencies given in a normal distribution, 19.5, instead of 33 per cent, exceed or fall short of the median accomplishment of the grade above or below.

All measures of overlapping thus far reported are incorrect as indicative of the numerical amount of overlapping in true ability. The correction in the measure of overlapping here derived is very easy to apply; a table of the frequencies in a normal distribution and the coefficient of reliability being the only new data required.

It is seen that great overlapping is as much an indictment of the test as it is the adequacy of the classification, unless it has been determined that the test itself is highly reliable. Three factors make for overlapping: (a) unreliability of tests; (b) unimportance of function tested; and (c) failure to properly classify, and only in so far as the first two are known not to be the causes can improper classification be established.

TRUMAN L. KELLEY.

Teachers College, Columbia University.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except June to August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Board of Editors

WILLIAM CHANDLER BAGLEY, CARL E SEASHORE,
J. CARLETON BELL, Managing Editor, GUY MONTROSE WHIPPLE.

COLLABORATORS

Sydney Alrutz,	H. H. Goddard,	Will S. Monroe,	L. William Stern
J. R. Angell,	Joseph Jastrow	E. C. Moore,	G. M. Stratton,
J. W. Baird,	G. E. Johnson,	M. V. O'Shea,	G. D. Strayer,
Bird T. Baldwin,	E. A. Kirkpatrick,	W. B. Pillsbury,	E. L. Thorndike,
W. Van Dyke Bingham,	E. H. Lindley,	S. H. Rowe,	E. B. Tichener,
E. F. Buchner,	Anna J. McKeag,	Sante de Sanctis,	J. B. Watson,
W. H. Burnham,	D. P. MacMillan,	E. G. Sanford,	W. H. Winch,
Ed. Claparede,	Adolf Meyer,	W. S. Small,	H. K. Wolfe,
S. S. Colvin,	J. B. Miner,	E. D. Starbuck,	A. H. Yoder.
W. F. Dearborn,			

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 per year of nine monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X APRIL, 1919 NUMBER 4

EDITORIAL

THE CLASSICS

Perhaps the war has not changed the fundamental aspects of human thinking as much as optimistic or pessimistic writers of fiction would have us believe, but it has certainly encouraged many people to take account of stock and has made them less tolerant of and complacent with traditional practices. There is a tendency to examine all values anew, and to inquire what each course of action contributes to the sum of human welfare. The failure of the Students' Army Training Corps movement (for it seems pretty generally acknowledged to have been a failure) may have been due to the impossible and unwarranted demands made upon the colleges, but it has afforded the critics of the old time methods and course of study, whether within or without the college, opportunity for renewed and effective attacks upon the uselessness and inanity of much traditional college work. These the colleges, made doubly sensitive by the increased cost of living and by depleted incomes, have keenly felt, and there are many indications of a distinct effort to meet popular demands. Even before the war there was a growing skepticism as to the educational value of

requiring Latin and Greek of all candidates for the bachelor of arts degree, and in most institutions Greek had already slipped into the position of an elective. Since the war this movement seems to have been accelerated, and the announcements of some of our most conservative colleges this spring show a noteworthy reduction in the requirements of the classics and mathematics. For example, Princeton has definitely abolished the requirement of Greek for either entrance or graduation, and has reduced the requirement of Latin to one year during the college course. Yale has made it possible to enter the university without any Latin at all, and has removed many of the advanced mathematics requirements for entrance to the Sheffield Scientific School. This moves an editorial writer in the *New York Times* to bewail the "notable advance of the movement for dehumanizing the humanities and deliberalizing the liberal arts," and leads the noted English classical scholar, Gilbert Murray, into an error of exaggeration. Writing to the *London Times*, he says: "A few months ago a strange and desolating portent occurred. All the American universities at one blow abolished Latin as a necessary subject for the arts degree. Presumably Latin will soon be as moribund as Greek in America."

While one is moved to smile at the idea of American universities being sufficiently united to do anything "at one blow," it is evident that the defenders of the vested rights of the classics to a monopoly of liberal education have reason to be alarmed. This alarm has resulted in the organization of the American Classical League and a plan for an aggressive campaign for the defense of classical learning. Let us hope that some of these champions of the classics may be induced by their ardor for the cause to make statistical and experimental studies on the processes involved in learning Latin and Greek, that the store of our knowledge in educational psychology may be enriched. Perhaps they may succeed in showing us what there is in the technical study of Latin vocabulary and sentence structure that is so remarkably liberalizing and humanizing. In the meantime let us press on in our efforts to work out in secondary schools and colleges a series of broadly liberal courses of study that will offer to each pupil the opportunity to come into contact with the biological, social and political history of the race (including the contributions of the Greeks and the Romans), and will give him training in applying this knowledge to contemporary problems.

J. C. B.

NOTES AND NEWS

Commenting on Mr. Maurice A. Garfinkel's study of *The Effect of the Summer Vacation on Ability in the Fundamentals of Arithmetic*, in the January issue of THIS JOURNAL, Dr. Arthur W. Kallom, of the department of educational investigation and measurement, Boston, writes that he has found among Boston pupils a similar deterioration in scores during the summer months. The differences between spring and fall scores are not so large, but a possible reason for this is the fact that the tests were given in April and October instead of in June and September. Tests given immediately before and immediately after the vacation would bring out more definitely any change due to that period than tests given several weeks before and after. Dr. Kallom deplores the use of the term "median variation" in Mr. Garfinkel's article for the difference between the June median score and the September median score. The term "median variation" has a definite significance in educational statistics, indicating the median of the variations of a group of scores from the central tendency of the group, and it only creates confusion when such a term is used in an obviously different and incorrect sense. In a new science, as that of educational measurements, it behooves us to be particularly careful in the use of our terminology in order to avoid ambiguity. So keenly is this need felt that the National Association of the Directors of Educational Research now has a committee at work on the standardization of statistical terms.

Dr. William H. Burnham, President of the Massachusetts Society for Mental Hygiene, has issued an interesting pamphlet stressing the need for an intensive study of the physical and mental characteristics of each child at the beginning of its school career. He suggests that the first year or two of school life should be devoted chiefly to tests of physiologic age; mental age; habits of observation, attention, association, and the like; general health conditions; and the development of the child's physical resources through play and other forms of motor activity. He hazards the view that there are two marked types of learners, one that learns easily by instruction, by hearing and seeing what others do, by definite directions and demonstration, that is, by the usual didactic methods of the school; and the other a type that learns chiefly by doing, and this by the child's own method. For the latter type the ordinary school procedure is deadly. It is of the utmost importance that these two types be distinguished as early in school life as possible, and the beginning of diagnosis should be at the entrance to school. While Dr. Burnham's assumption of sharply defined types does not seem to be in accord with the results of most experimental investigations, which show continuous variation in every trait studied, this in no way weakens the force of his appeal for a more intensive study of school entrants.

Dr. L. W. Sackett, of the University of Texas, Austin, Texas, has recently issued in neat pamphlet form the Bell-McCollum tests in American history. The eight-page pamphlet includes not only the tests themselves, but the norms derived from their original application, and a weighted evaluation of each question on the basis of those results. The tests may be obtained from Dr. Sackett at the rate of $1.50 per hundred, postage extra.

The annual meeting of the National Association of Directors of Educational Research was held at Chicago, February 27. The forenoon session was devoted exclusively to reports by the members on various technical researches in measuring intelligence and in reading, spelling, geography and composition. The afternoon program covered a more general field, centering about the need for, and the organization and support of educational research work. Dr. B. R. Buckinghan, director of the Bureau of Educational Research, University of Illinois, was elected president, and Dr. Ernest J. Ashbaugh, director educational extension service, University of Iowa, secretary-treasurer.

The March meeting of the New York Society for the Experimental Study of Education was devoted to the consideration of "Problems in the Education of Exceptional Children." The program consisted of two parts. In Part 1, *The Exceptionally Dull Child*, Miss Elizabeth E. Farrell, inspector of ungraded classes, discussed the "After-care of Ungraded Class Children," and Mr. George H. Chatfield, assistant director of attendance spoke on "Problems in Compulsory Education." In Part 11, *The Exceptionally Bright Child*, Miss Anna Gillingham, Ethical Culture School reported on "Intellectual Brightness and School Progress," and Principal Louis A. Marks, Public School 64, Manhattan, spoke on "Intelligence Tests as a Basis for School Grading." The discussion was led by Dr. John L. Tildsley, associate superintendent in charge of high schools, and Professor James E. Lough, of New York University.